The Current State of Data and Model Transparency

Aaron Y. Lee MD MSCI
University of Washington
Associate Professor
C. Dan and Irene Hunter Endowed Professor

Computational Ophthalmology

Disclosures

- NVIDIA Corporation
- Microsoft Corporation
- Amazon
- Boehringer Ingelheim
- Novartis
- Genentech / Roche
- Santen
- Johnson and Johnson
- Gyroscope
- Carl Zeiss Meditec
- Topcon
- iCareWorld
- Heidelberg
- Optomed

Bhavesh Patel

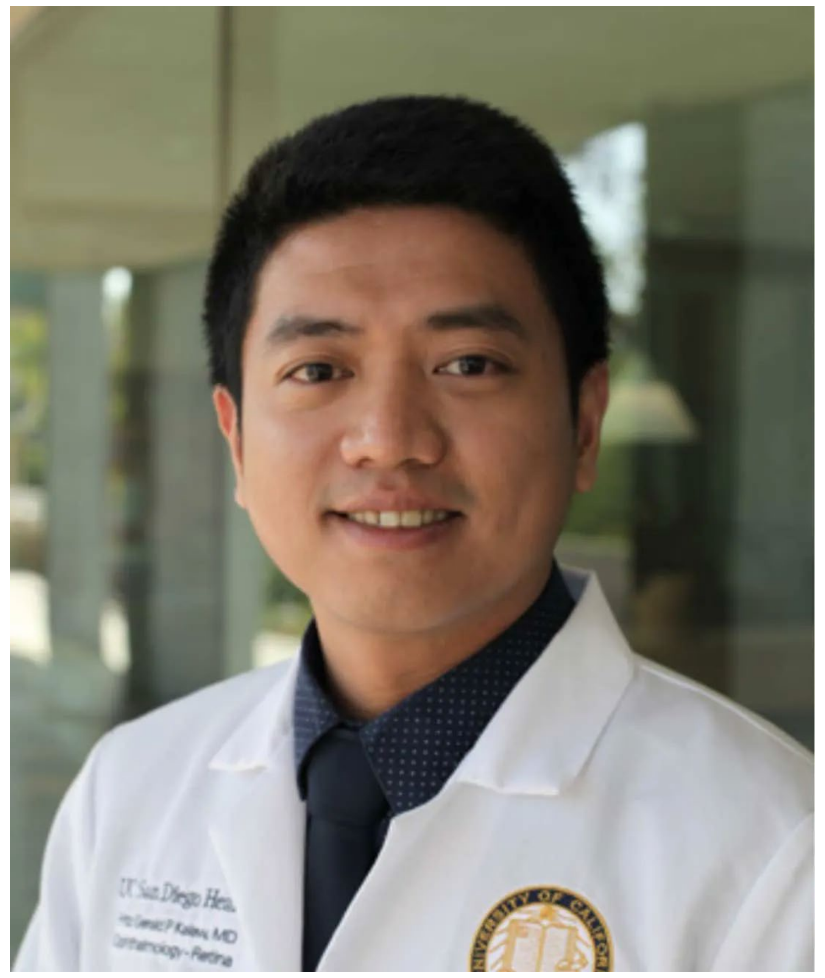
Team

AI-READI

AI Ready and Equitable Atlas for Diabetes Insights

Anna Heinke

Lingling Huang



Fritz Kalaw

Kyongmi Simpkins

● DATA FOCUSED

- Data Sheets ● ● ● ● ● ● ●
- Data Statements ● ● ● ● ●
- Data Nutrition Labels ● ● ● ● ●
- Data Cards for NLP ● ● ● ● ●
- Dataset Development Lifecycle Documentation Framework ● ● ● ● ● ● ●
- Data Cards ● ● ● ● ●

● MODELS & METHODS FOCUSED

- Model Cards ● ● ● ● ● ● ●
- Value Cards ● ● ● ● ●
- Method Cards ● ●
- Consumer Labels for Models ● ● ● ●

● SYSTEMS FOCUSED

- System Cards ● ● ● ● ●
- FactSheets ● ● ● ●
- ABOUT ML ● ● ● ● ● ● ●

SAMPLE OF POTENTIAL AUDIENCES

● ML Engineers ● Model Developers/Reviewers ● Students ● Policymakers

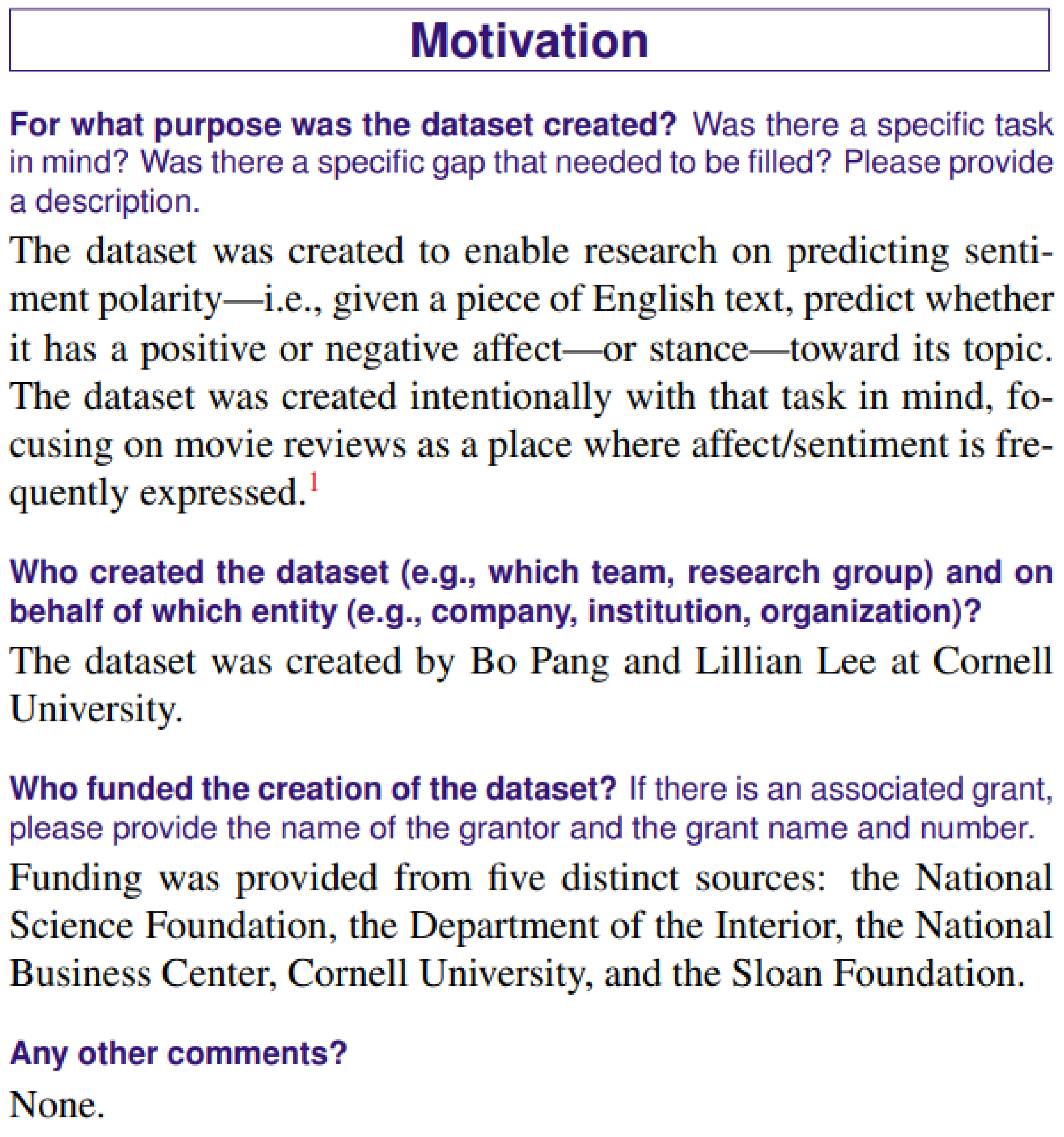
● Ethicists ● Data Scientists/Business Analysts ● Impacted Individuals

https://huggingface.co/blog/model-cards

Background

Datasheets are a popularly suggested metadata file…
but there are many variants

Datasheet



Motivation

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

The dataset was created to enable research on predicting sentiment polarity—i.e., given a piece of English text, predict whether it has a positive or negative affect—or stance—toward its topic. The dataset was created intentionally with that task in mind, focusing on movie reviews as a place where affect/sentiment is frequently expressed.[1]

Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

The dataset was created by Bo Pang and Lillian Lee at Cornell University.

Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.

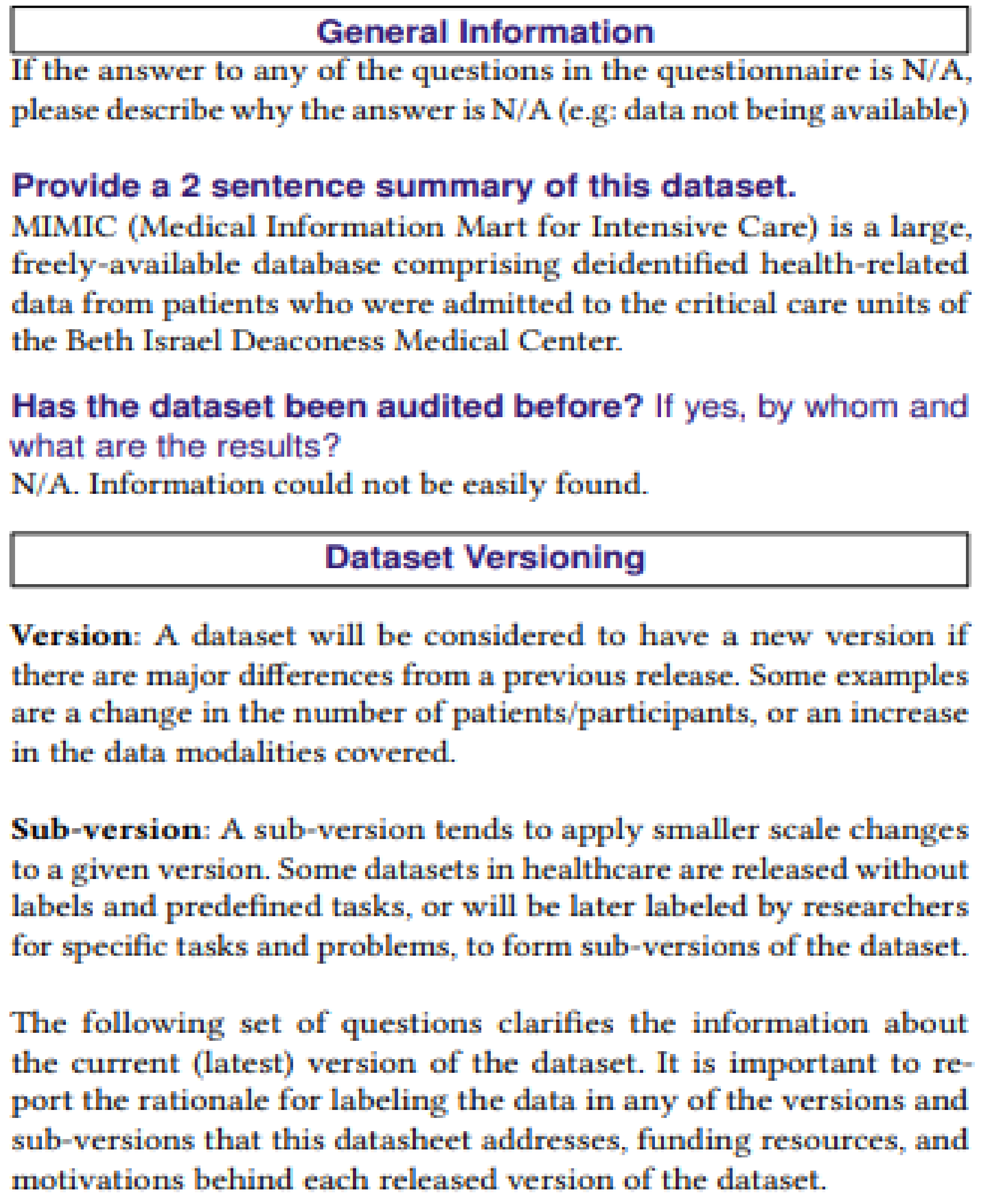
Funding was provided from five distinct sources: the National Science Foundation, the Department of the Interior, the National Business Center, Cornell University, and the Sloan Foundation.

Any other comments?

None.

VS

Healthsheet



General Information

If the answer to any of the questions in the questionnaire is N/A, please describe why the answer is N/A (e.g: data not being available)

Provide a 2 sentence summary of this dataset.

MIMIC (Medical Information Mart for Intensive Care) is a large, freely-available database comprising deidentified health-related data from patients who were admitted to the critical care units of the Beth Israel Deaconess Medical Center.

Has the dataset been audited before? If yes, by whom and what are the results?

N/A. Information could not be easily found.

Dataset Versioning

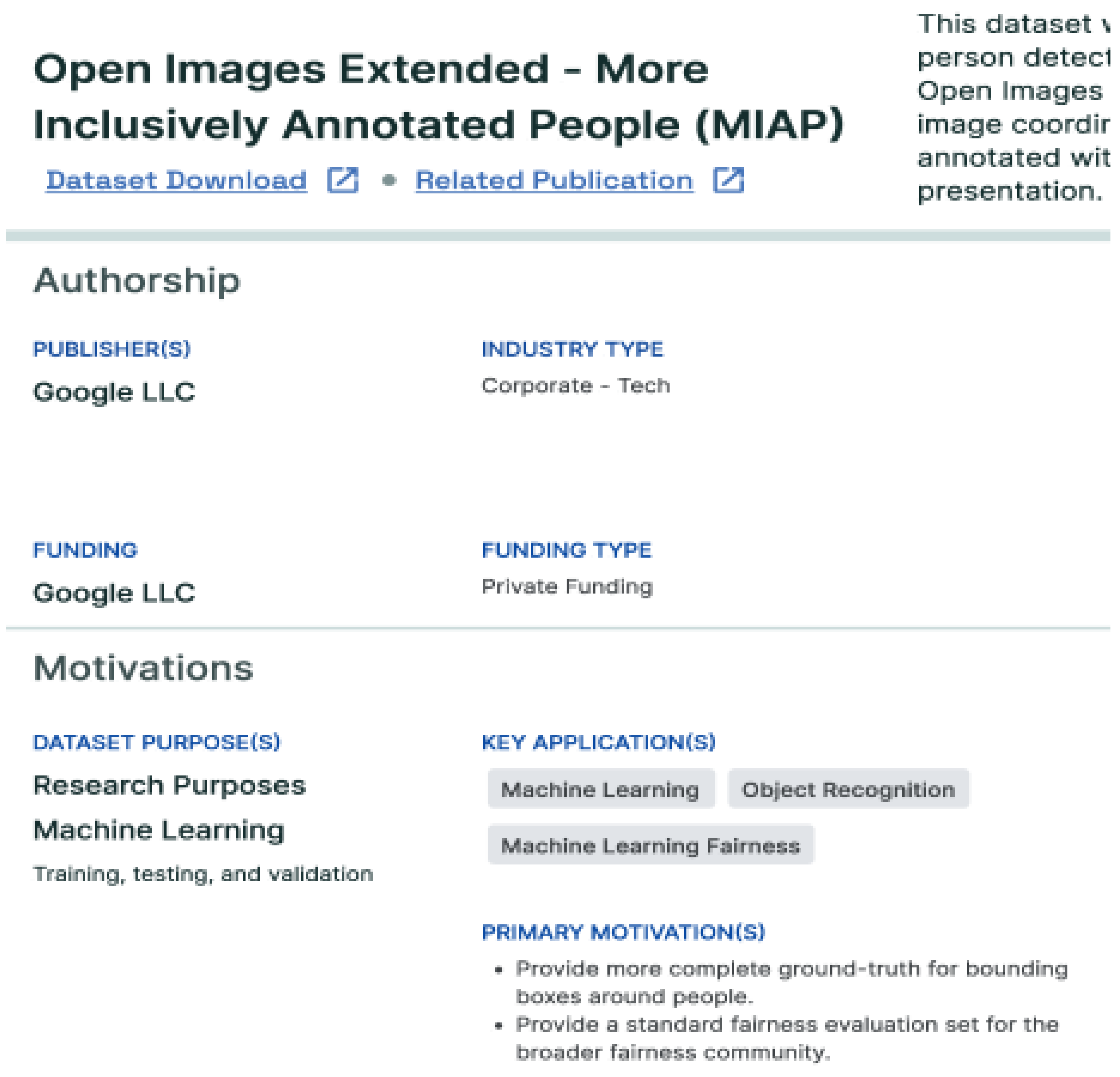
Version: A dataset will be considered to have a new version if there are major differences from a previous release. Some examples are a change in the number of patients/participants, or an increase in the data modalities covered.

Sub-version: A sub-version tends to apply smaller scale changes to a given version. Some datasets in healthcare are released without labels and predefined tasks, or will be later labeled by researchers for specific tasks and problems, to form sub-versions of the dataset.

The following set of questions clarifies the information about the current (latest) version of the dataset. It is important to report the rationale for labeling the data in any of the versions and sub-versions that this datasheet addresses, funding resources, and motivations behind each released version of the dataset.

VS

Data card



Open Images Extended - More Inclusively Annotated People (MIAP)

Dataset Download • Related Publication

This dataset v
person detect
Open Images
image coordir
annotated wit
presentation.

Authorship

PUBLISHER(S)	INDUSTRY TYPE
Google LLC	Corporate - Tech

FUNDING	FUNDING TYPE
Google LLC	Private Funding

Motivations

DATASET PURPOSE(S)

Research Purposes

Machine Learning

Training, testing, and validation

KEY APPLICATION(S)

Machine Learning · Object Recognition · Machine Learning Fairness

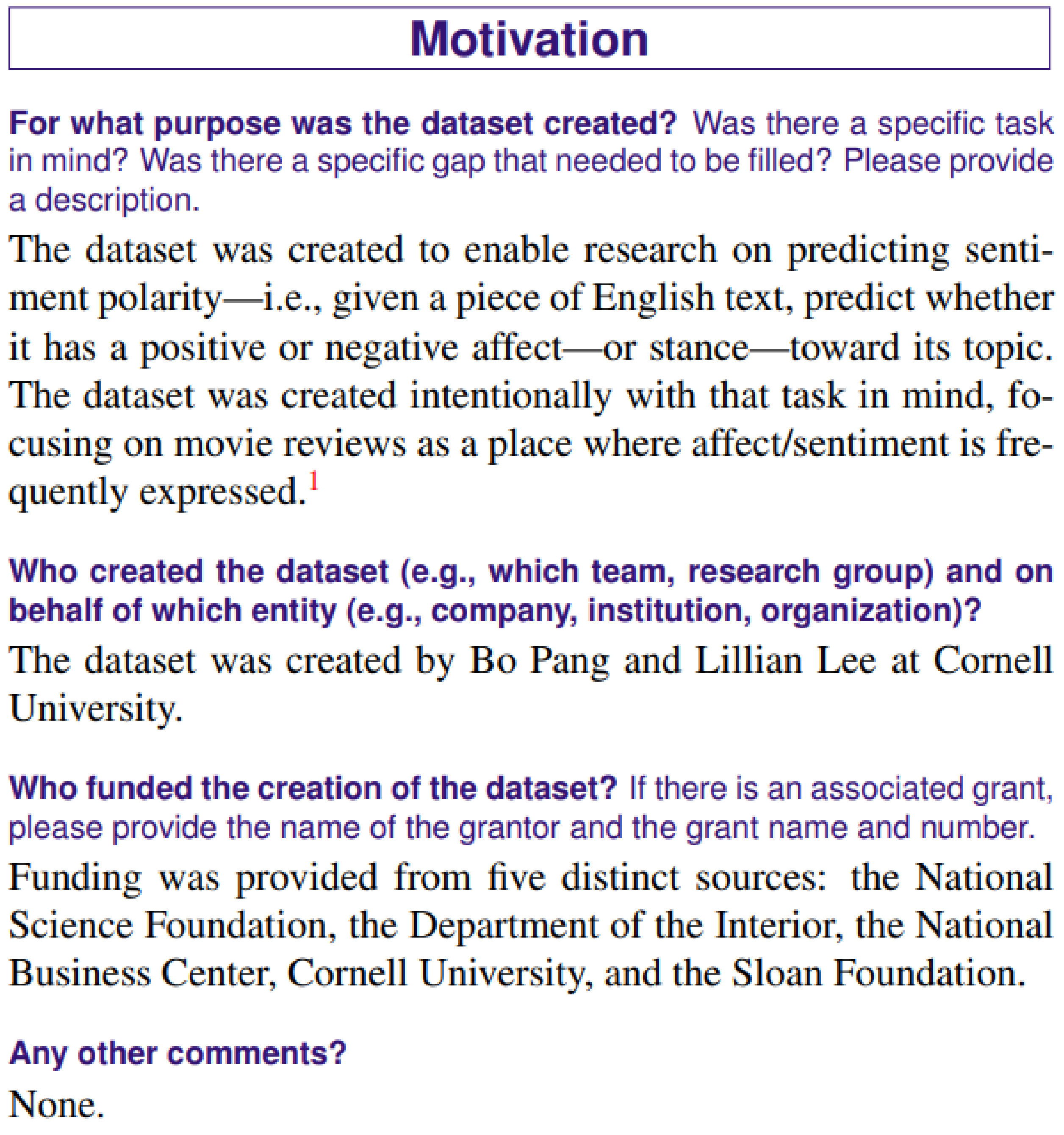
PRIMARY MOTIVATION(S)

- Provide more complete ground-truth for bounding boxes around people.
- Provide a standard fairness evaluation set for the broader fairness community.

Datasheet

Datasheet
What is it?

A datasheet is a document consisting of a series of questions/answers that is intended to document motivation, composition, collection process, recommended uses, and so on for a dataset



Motivation

For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

The dataset was created to enable research on predicting sentiment polarity—i.e., given a piece of English text, predict whether it has a positive or negative affect—or stance—toward its topic. The dataset was created intentionally with that task in mind, focusing on movie reviews as a place where affect/sentiment is frequently expressed.[1]

Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

The dataset was created by Bo Pang and Lillian Lee at Cornell University.

Who funded the creation of the dataset? If there is an associated grant, please provide the name of the grantor and the grant name and number.

Funding was provided from five distinct sources: the National Science Foundation, the Department of the Interior, the National Business Center, Cornell University, and the Sloan Foundation.

Any other comments?

None.

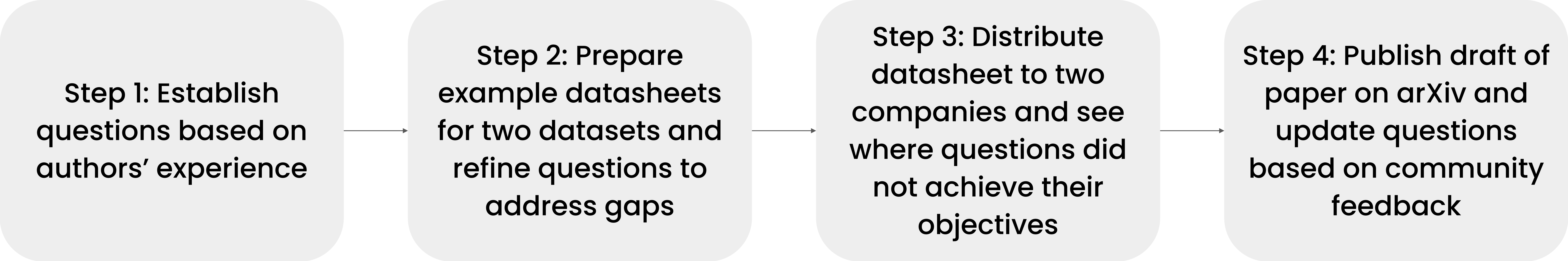
Datasheet

Timeline

Start: ?	It is not specified when this effort was started
March 2018	Preprint v1 published on arXiv https://doi.org/10.48550/arXiv.1803.09010
2018-2021	Several revised versions published on arXiv
December 2021	Final version published in Communications of the ACM https://doi.org/10.1145/3458723

Datasheet

How was it developed?



Step 1: Establish questions based on authors' experience

Step 2: Prepare example datasheets for two datasets and refine questions to address gaps

Step 3: Distribute datasheet to two companies and see where questions did not achieve their objectives

Step 4: Publish draft of paper on arXiv and update questions based on community feedback

Datasheet
How it is structured?

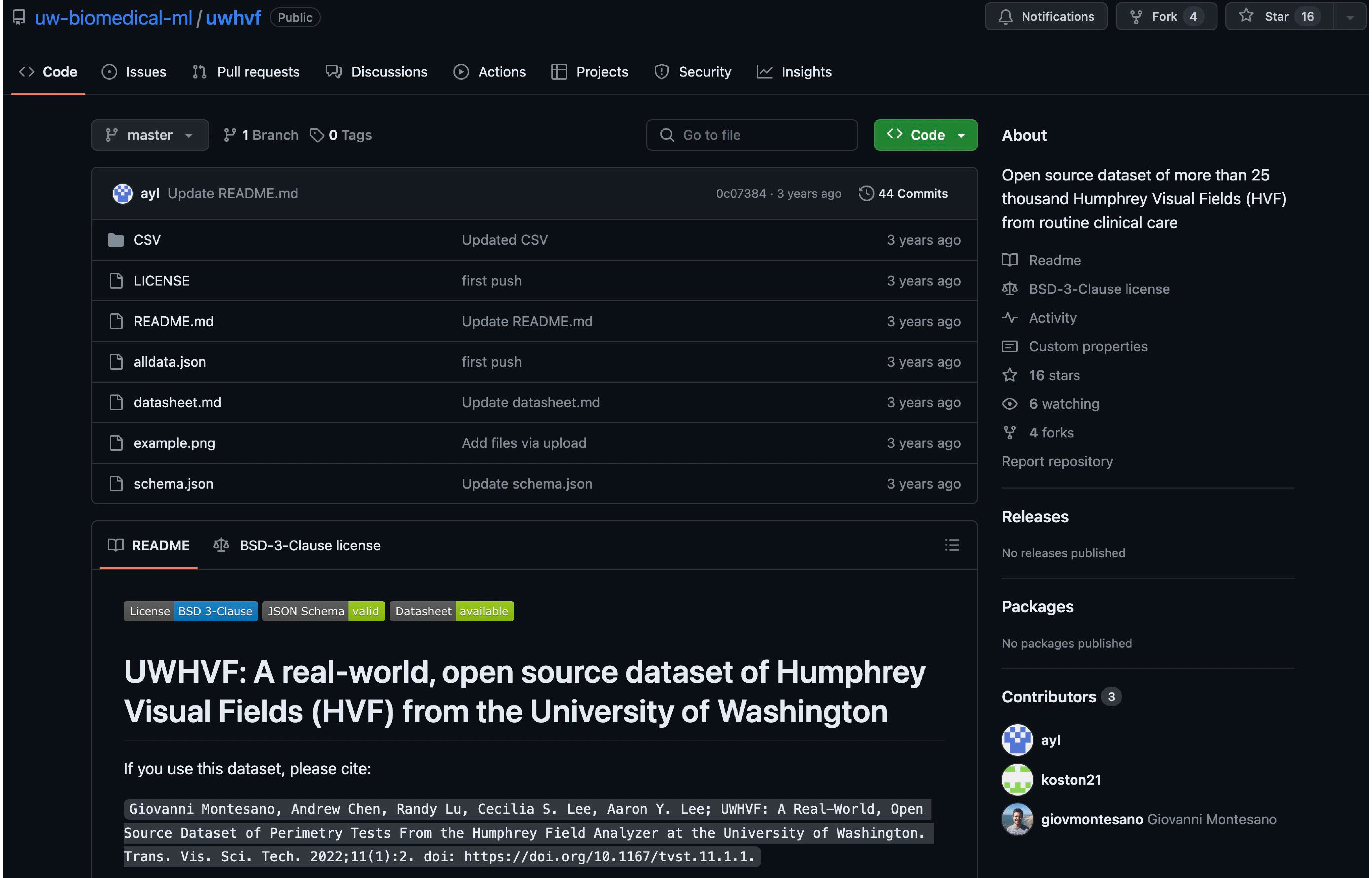
- 7 sections and 56 questions
  1. “**Motivation**”: Reasons for creating the dataset, funding source, etc. 4 questions
  2. “**Composition**” : Describe the content of the dataset, de-identification level, etc. 16 questions
  3. “**Collection Process**” : Describe the data collection process. 12 questions
  4. “**Preprocessing/cleaning/labeling**” : Describe data processing. 4 questions
  5. “**Uses”** : Specify tasks for which the dataset should and should not be used. 6 questions
  6. **“Distribution”**: Describe the dataset distribution/sharing process. 7 questions
  7. “**Maintenance”:** Communicate plan for maintaining the dataset. 7 questions

Datasheet

Are there templates/tools available to create it?

- Note that the paper mentions: “We emphasize that the process of creating a datasheet is not intended to be automated. Although automated documentation processes are convenient, they run counter to our objective of encouraging dataset creators to carefully reflect on the process of creating, distributing, and maintaining a dataset.”
- We could not find any tool that helps preparing a datasheet.
- Templates are available in different formats:
  - Markdown: https://github.com/fau-masters-collected-works-cgarbin/datasheet-for-dataset-template
  - Markdown: https://github.com/JRMeyer/markdown-datasheet-for-datasets/blob/master/DATASHEET.md
  - JSON: https://github.com/JRMeyer/json-datasheet-for-datasets/blob/main/DATASHEET.json
  - LaTex: https://github.com/AudreyBeard/Datasheets-for-Datasets-Template
  - Latex: https://www.overleaf.com/latex/templates/datasheet-for-dataset-template/jgqyyzyprxth

    https://www.overleaf.com/project/6581f30f780f8c448b45ea02



ayl Update README.md 0c07384 · 3 years ago 44 Commits

CSV	Updated CSV	3 years ago
LICENSE	first push	3 years ago
README.md	Update README.md	3 years ago
alldata.json	first push	3 years ago
datasheet.md	Update datasheet.md	3 years ago
example.png	Add files via upload	3 years ago
schema.json	Update schema.json	3 years ago

License BSD 3-Clause JSON Schema valid Datasheet available

UWHVF: A real-world, open source dataset of Humphrey Visual Fields (HVF) from the University of Washington

If you use this dataset, please cite:

`Giovanni Montesano, Andrew Chen, Randy Lu, Cecilia S. Lee, Aaron Y. Lee; UWHVF: A Real-World, Open Source Dataset of Perimetry Tests From the Humphrey Field Analyzer at the University of Washington. Trans. Vis. Sci. Tech. 2022;11(1):2. doi: https://doi.org/10.1167/tvst.11.1.1.`

About

Open source dataset of more than 25 thousand Humphrey Visual Fields (HVF) from routine clinical care

- Readme
- BSD-3-Clause license
- Activity
- Custom properties
- 16 stars
- 6 watching
- 4 forks

Report repository

Releases

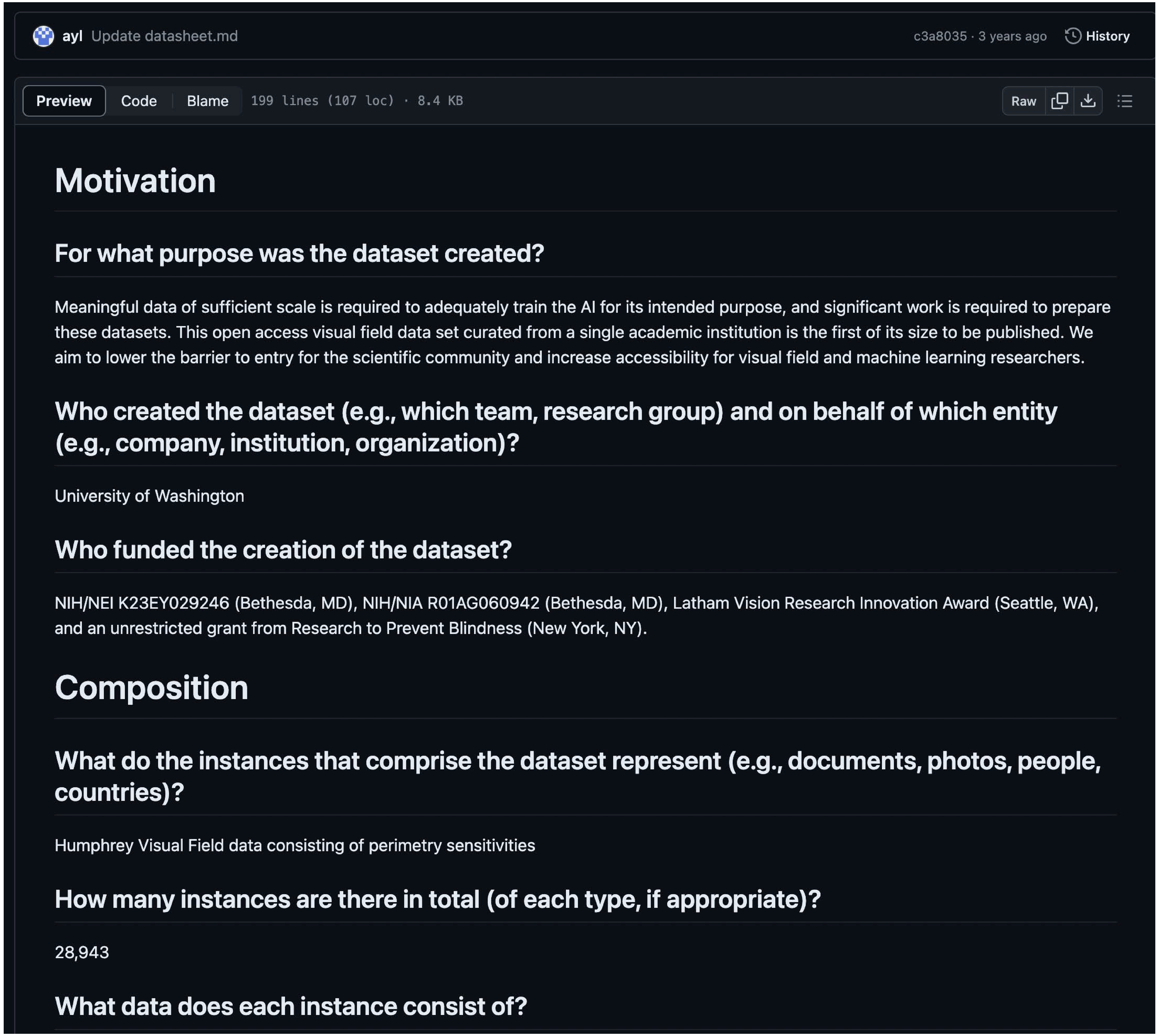
No releases published

Packages

No packages published

Contributors 3

- ayl
- koston21
- giovmontesano Giovanni Montesano



Motivation

For what purpose was the dataset created?

Meaningful data of sufficient scale is required to adequately train the AI for its intended purpose, and significant work is required to prepare these datasets. This open access visual field data set curated from a single academic institution is the first of its size to be published. We aim to lower the barrier to entry for the scientific community and increase accessibility for visual field and machine learning researchers.

Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

University of Washington

Who funded the creation of the dataset?

NIH/NEI K23EY029246 (Bethesda, MD), NIH/NIA R01AG060942 (Bethesda, MD), Latham Vision Research Innovation Award (Seattle, WA), and an unrestricted grant from Research to Prevent Blindness (New York, NY).

Composition

What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?

Humphrey Visual Field data consisting of perimetry sensitivities

How many instances are there in total (of each type, if appropriate)?

28,943

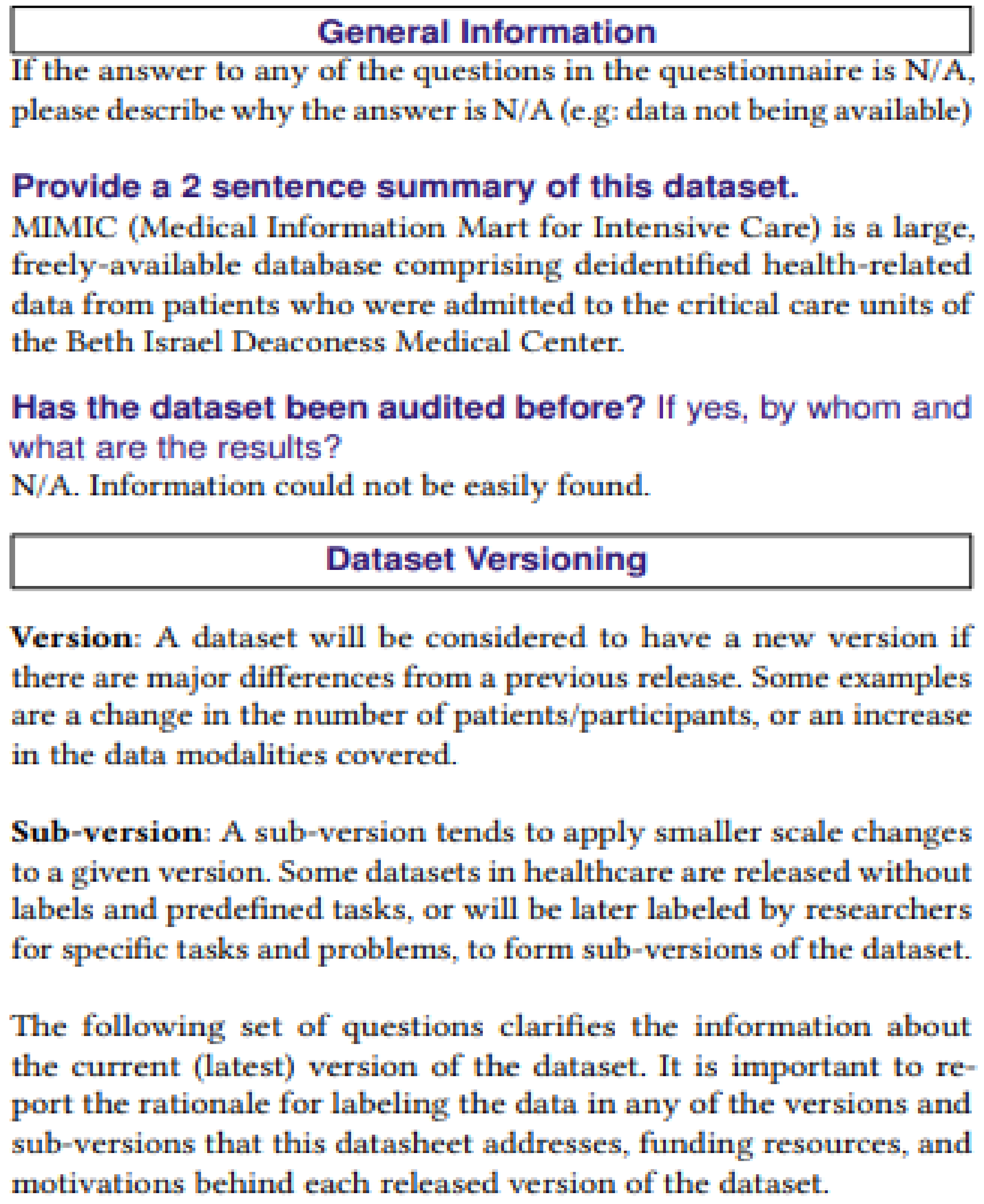
What data does each instance consist of?

Healthsheet

Healthsheet

What is it?

Healthsheet is a contextualized adaptation of the original datasheet questionnaire for health specific applications.



General Information

If the answer to any of the questions in the questionnaire is N/A, please describe why the answer is N/A (e.g: data not being available)

Provide a 2 sentence summary of this dataset.
MIMIC (Medical Information Mart for Intensive Care) is a large, freely-available database comprising deidentified health-related data from patients who were admitted to the critical care units of the Beth Israel Deaconess Medical Center.

Has the dataset been audited before? If yes, by whom and what are the results?
N/A. Information could not be easily found.

Dataset Versioning

Version: A dataset will be considered to have a new version if there are major differences from a previous release. Some examples are a change in the number of patients/participants, or an increase in the data modalities covered.

Sub-version: A sub-version tends to apply smaller scale changes to a given version. Some datasets in healthcare are released without labels and predefined tasks, or will be later labeled by researchers for specific tasks and problems, to form sub-versions of the dataset.

The following set of questions clarifies the information about the current (latest) version of the dataset. It is important to report the rationale for labeling the data in any of the versions and sub-versions that this datasheet addresses, funding resources, and motivations behind each released version of the dataset.

Healthsheet
How does it differ from datasheet?

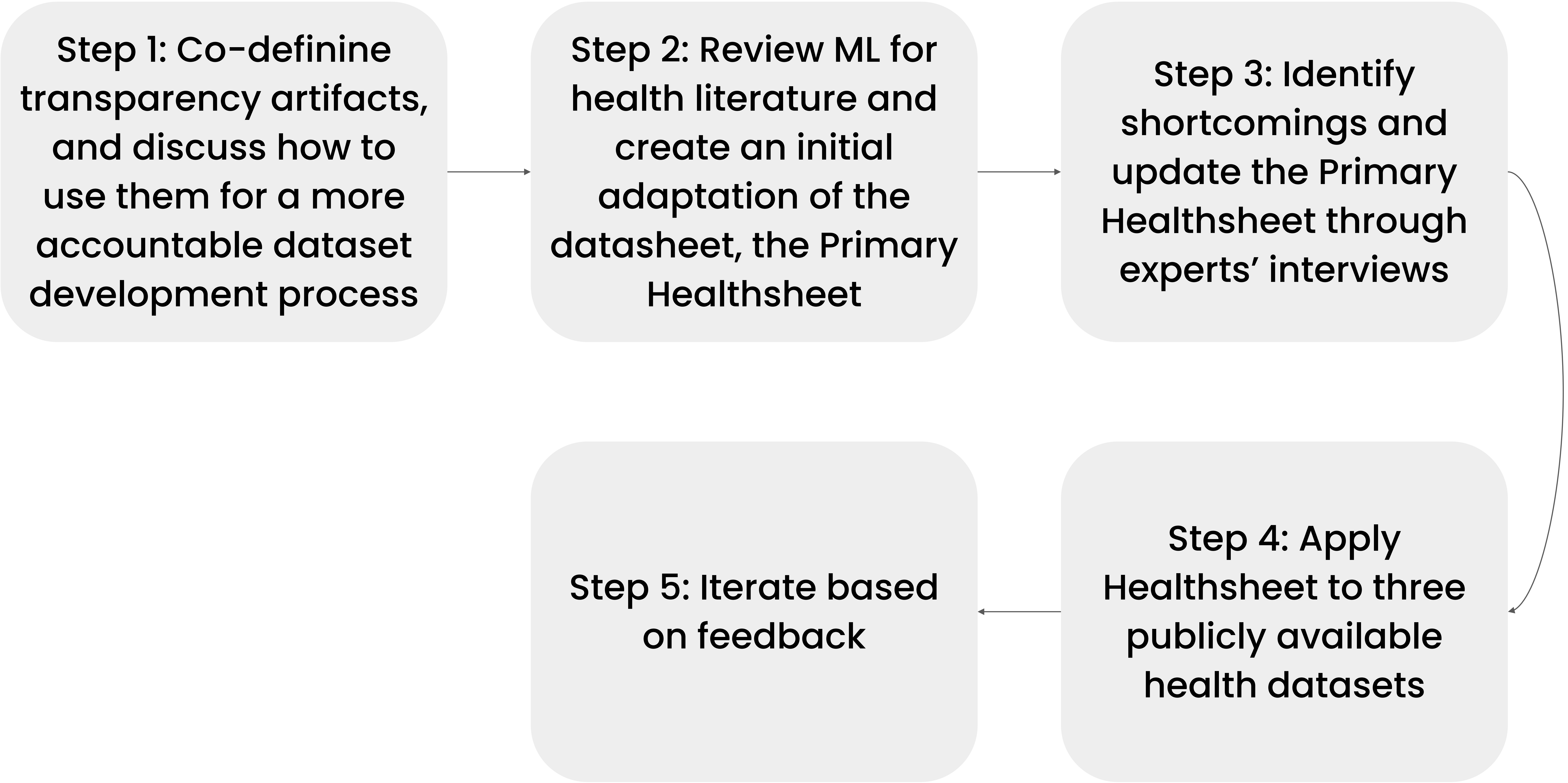
	Healthsheet	**Datasheet**
Purpose and Focus	Tailored for healthcare datasets	Primarily designed for machine learning datasets
Context and industry	Targeted at healthcare industry	Applicable across various industries using ML
Elements and sections	- Dataset versioning- Accessibility- Demographic info- Racism/social conditions	- Motivation- Composition- Collection process- Fairness considerations
Use cases	Clinical research, healthcare applications.	Machine learning research, model development.
Interdisciplinary collaboration	Collaboration with healthcare professionals, ethicists.	Collaboration between data scientists and domain experts.
Depth of information	Detailed information on demographic factors, accessibility.	In-depth insights into dataset creation, biases.
Application scope	Clinical research, healthcare analytics.	General machine learning applications.

Healthsheet
Timeline

Start: ?	The starting date of this effort is not specified.
Feb 2022	Publication of the associated paper on arXiv https://doi.org/10.48550/arXiv.2202.13028
June 2022	Publication of the associated paper in Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency https://doi.org/10.1145/3531146.3533239

Healthsheet

How was it developed?



Step 1: Co-definine transparency artifacts, and discuss how to use them for a more accountable dataset development process

Step 2: Review ML for health literature and create an initial adaptation of the datasheet, the Primary Healthsheet

Step 3: Identify shortcomings and update the Primary Healthsheet through experts' interviews

Step 4: Apply Healthsheet to three publicly available health datasets

Step 5: Iterate based on feedback

Healthsheet

How it is structured?

- General Information
- Dataset versioning
- Motivation
- Data composition
- Collection and use of demographic information
- Pre-processing, de-identification
- Labeling and subjectivity of labeling
- Collection process
- Uses
- Data distribution
- Maintenance

Healthsheet for “Development of An Open-Source Annotated Glaucoma Medication Dataset from Clinical Notes in the Electronic Health Record”

Jimmy S. Chen, MD; Wei-Chun Lin, MD; Sen Yang, MD; Michael F. Chiang, MD, MA; Michelle R. Hribar, PhD

General Information

If the answer to any of the questions in the questionnaire is N/A, please describe why the answer is N/A (e.g: data not being available)

Provide a 2 sentence summary of this dataset.

This dataset consists of clinical notes for glaucoma patients at OHSU seen over 2019. These notes were de-identified for protected health information (PHI) and annotated for glaucoma medications.

Has the dataset been audited before? If yes, by whom and what are the results?

No, this dataset has never been previously audited.

Dataset Versioning

Version: A dataset will be considered to have a new version if there are major differences from a previous release. Some examples are a change in the number of patients/participants, or an increase in the data modalities covered.

Subversion: A sub-version tends to apply smaller scale changes to a given version. Some datasets in healthcare are released without labels and predefined tasks, or will be later labeled by researchers for specific tasks and problems, to form sub-versions of the dataset. for labeling the data in any of the versions and sub-versions that this datasheet addresses, funding resources, and motivations behind each released version of the dataset.

Does the dataset get released as static versions or is it dynamically updated?
a. If static, how many versions of the dataset exist?
b.If dynamic, how frequently is the dataset updated?

This dataset will be static, with updates reserved for errata.

Is this datasheet created for the original version of the dataset? If not, which version of the dataset is this datasheet for?

This datasheet was created for the original version of the dataset (1.0).

Are there any datasheets created for any versions of this dataset?

No other prior datasheets or prior versions of this dataset exist.

Does the current version/subversion of the dataset come with predefined task(s), labels, and recommended data splits (e.g., for training, development/validation, testing)? If yes, please provide a high-level description of the introduced tasks, data splits, and labeling, and explain the rationale behind them. Please provide the related links and references. If not, is there any resource (website, portal, etc.) to keep track of all defined tasks and/or label definitions?

Annotated glaucoma medications are included in this dataset. No splits for training, validation, or testing are included in this dataset.

If the dataset has multiple versions, and this datasheet represents one of them, answer

Healthsheet

Examples of datasets using it

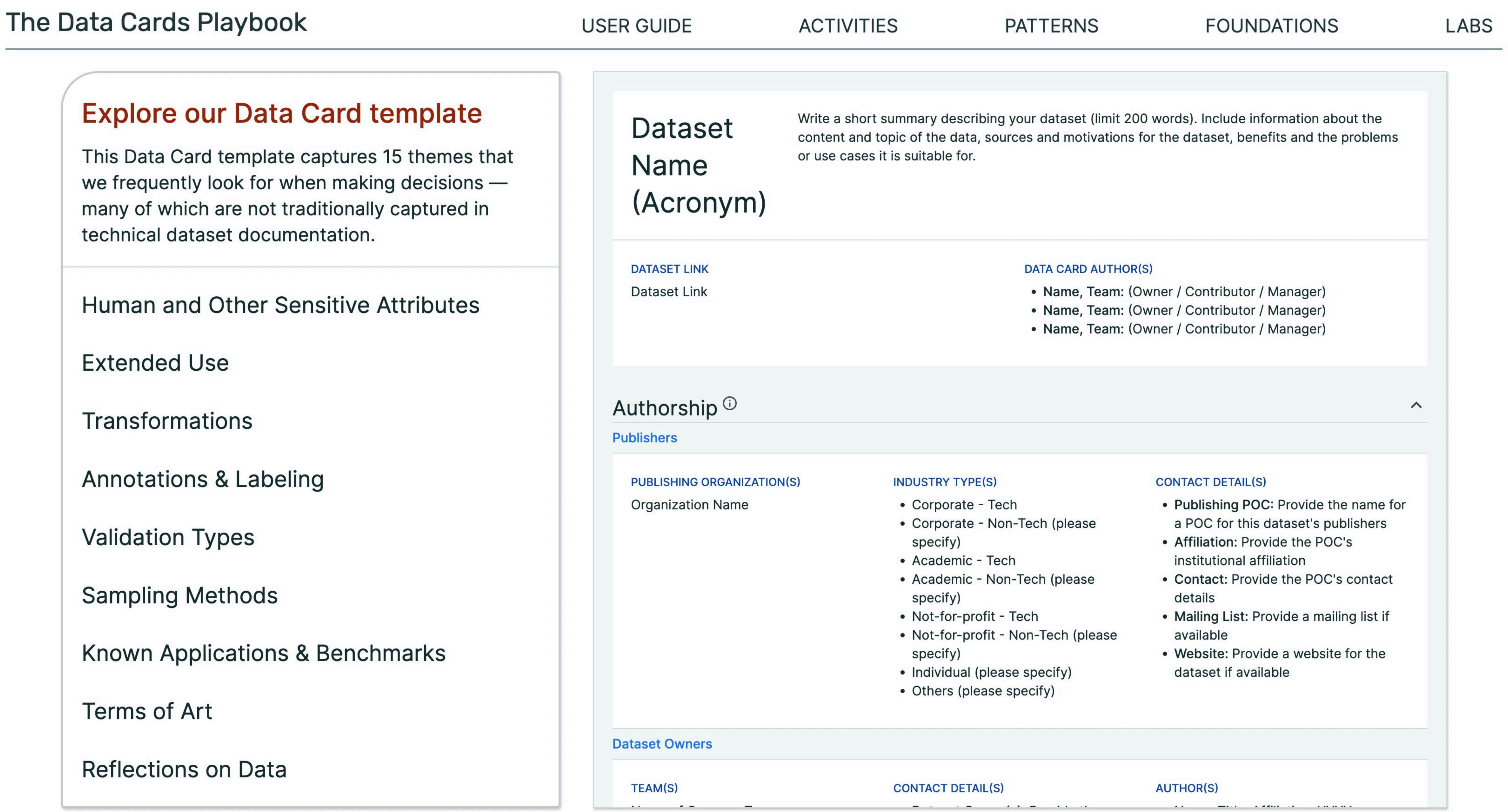
- Open Dataset of Flat-mounted Images for the Oxygen-induced Retinopathy Mouse Model: https://doi.org/10.6084/m9.figshare.23690973.v3

Other efforts

Data Cards: Purposeful and Transparent Dataset Documentation for Responsible AI

Mahima Pushkarna, Google Research, Canada, mahimap@google.com
Andrew Zaldivar, Google Research, USA, andrewzaldivar@google.com
Oddur Kjartansson, Google Research, United Kingdom, oddur@google.com

DOI: https://doi.org/10.1145/3531146.3533231
FAccT '22: 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, June 2022



The Data Cards Playbook

Explore our Data Card template

This Data Card template captures 15 themes that we frequently look for when making decisions — many of which are not traditionally captured in technical dataset documentation.

Human and Other Sensitive Attributes

Extended Use

Transformations

Annotations & Labeling

Validation Types

Sampling Methods

Known Applications & Benchmarks

Terms of Art

Reflections on Data

Dataset Name (Acronym)

Write a short summary describing your dataset (limit 200 words). Include information about the content and topic of the data, sources and motivations for the dataset, benefits and the problems or use cases it is suitable for.

DATASET LINK
Dataset Link

DATA CARD AUTHOR(S)
- **Name, Team:** (Owner / Contributor / Manager)
- **Name, Team:** (Owner / Contributor / Manager)
- **Name, Team:** (Owner / Contributor / Manager)

Authorship ⓘ

Publishers

PUBLISHING ORGANIZATION(S)
Organization Name

INDUSTRY TYPE(S)
- Corporate - Tech
- Corporate - Non-Tech (please specify)
- Academic - Tech
- Academic - Non-Tech (please specify)
- Not-for-profit - Tech
- Not-for-profit - Non-Tech (please specify)
- Individual (please specify)
- Others (please specify)

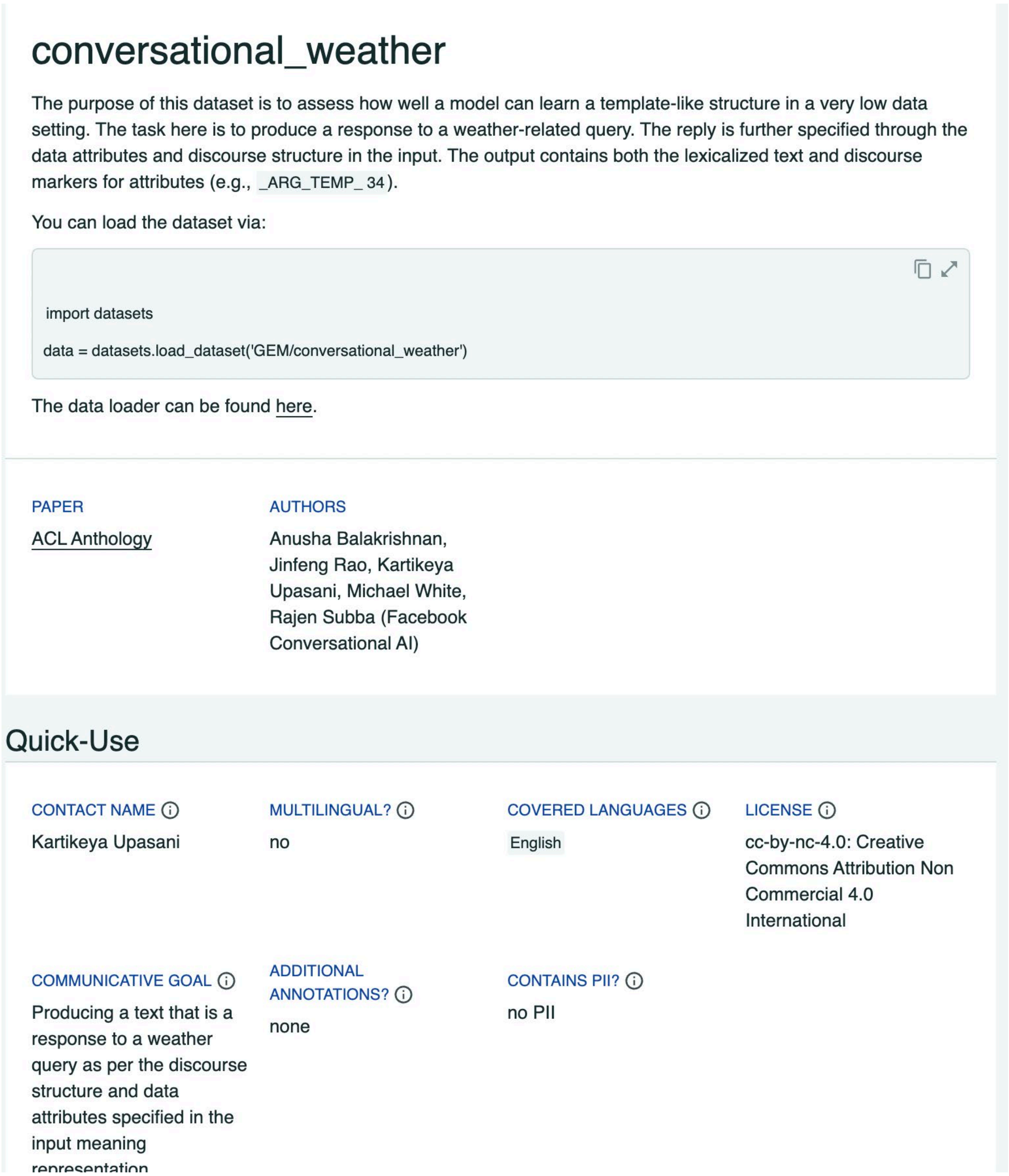
CONTACT DETAIL(S)
- **Publishing POC:** Provide the name for a POC for this dataset's publishers
- **Affiliation:** Provide the POC's institutional affiliation
- **Contact:** Provide the POC's contact details
- **Mailing List:** Provide a mailing list if available
- **Website:** Provide a website for the dataset if available

Dataset Owners

TEAM(S)

CONTACT DETAIL(S)

AUTHOR(S)



conversational_weather

The purpose of this dataset is to assess how well a model can learn a template-like structure in a very low data setting. The task here is to produce a response to a weather-related query. The reply is further specified through the data attributes and discourse structure in the input. The output contains both the lexicalized text and discourse markers for attributes (e.g., _ARG_TEMP_ 34).

You can load the dataset via:

```
import datasets

data = datasets.load_dataset('GEM/conversational_weather')
```

The data loader can be found here.

PAPER
ACL Anthology

AUTHORS
Anusha Balakrishnan, Jinfeng Rao, Kartikeya Upasani, Michael White, Rajen Subba (Facebook Conversational AI)

Quick-Use

CONTACT NAME ⓘ
Kartikeya Upasani

MULTILINGUAL? ⓘ
no

COVERED LANGUAGES ⓘ
English

LICENSE ⓘ
cc-by-nc-4.0: Creative Commons Attribution Non Commercial 4.0 International

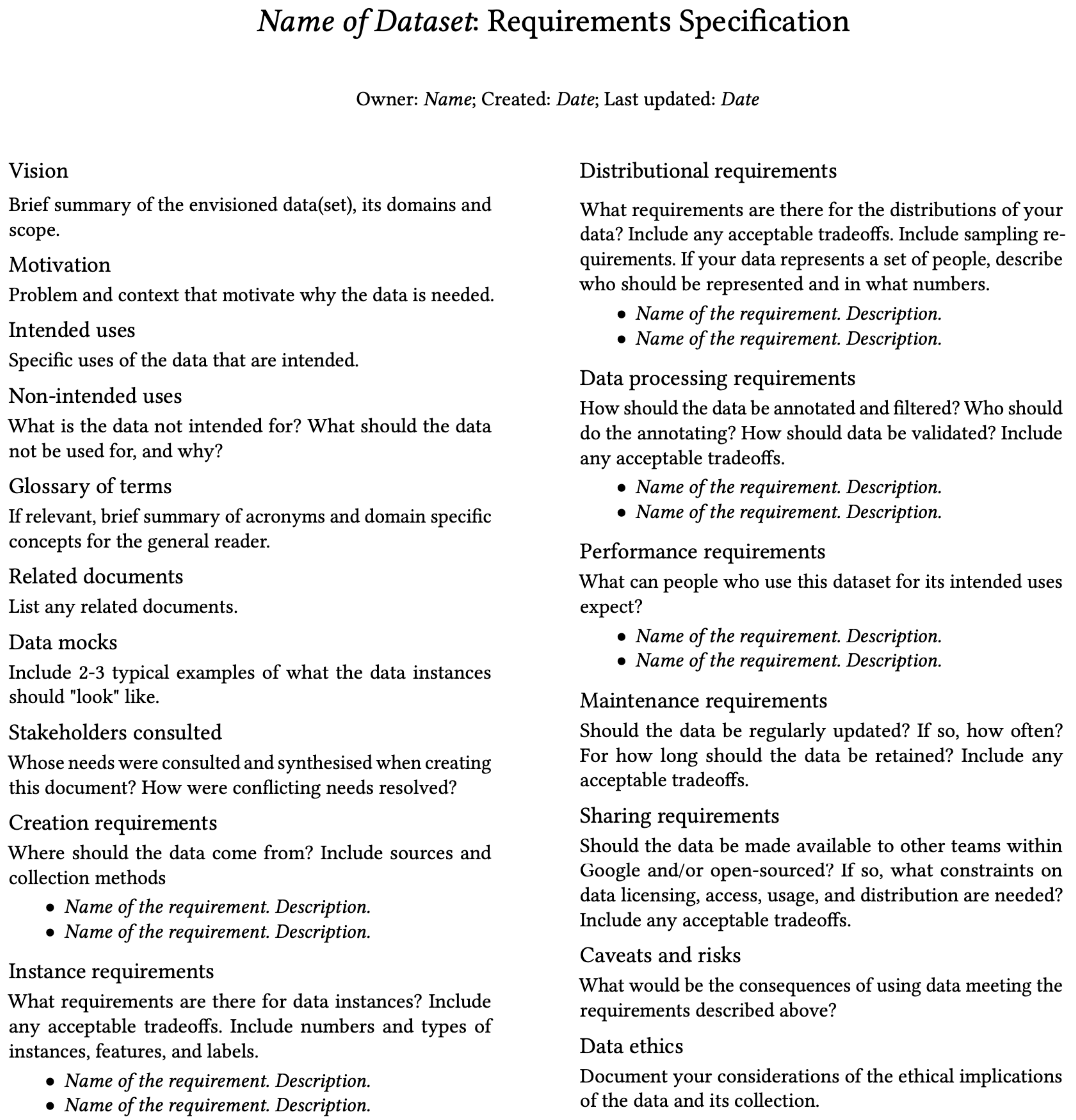
COMMUNICATIVE GOAL ⓘ
Producing a text that is a response to a weather query as per the discourse structure and data attributes specified in the input meaning representation

ADDITIONAL ANNOTATIONS? ⓘ
none

CONTAINS PII? ⓘ
no PII

Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure

Ben Hutchinson, Andrew Smart, Alex Hanna, Emily Denton, Christina Greer, Oddur Kjartansson, Parker Barnes, Margaret Mitchell
{benhutch,andrewsmart,alexhanna,dentone,ckuhn,oddur,parkerbarnes,mmitchellai}@google.com



Name of Dataset: Requirements Specification

Owner: *Name*; Created: *Date*; Last updated: *Date*

Vision

Brief summary of the envisioned data(set), its domains and scope.

Motivation

Problem and context that motivate why the data is needed.

Intended uses

Specific uses of the data that are intended.

Non-intended uses

What is the data not intended for? What should the data not be used for, and why?

Glossary of terms

If relevant, brief summary of acronyms and domain specific concepts for the general reader.

Related documents

List any related documents.

Data mocks

Include 2-3 typical examples of what the data instances should "look" like.

Stakeholders consulted

Whose needs were consulted and synthesised when creating this document? How were conflicting needs resolved?

Creation requirements

Where should the data come from? Include sources and collection methods

- *Name of the requirement. Description.*
- *Name of the requirement. Description.*

Instance requirements

What requirements are there for data instances? Include any acceptable tradeoffs. Include numbers and types of instances, features, and labels.

- *Name of the requirement. Description.*
- *Name of the requirement. Description.*

Distributional requirements

What requirements are there for the distributions of your data? Include any acceptable tradeoffs. Include sampling requirements. If your data represents a set of people, describe who should be represented and in what numbers.

- *Name of the requirement. Description.*
- *Name of the requirement. Description.*

Data processing requirements

How should the data be annotated and filtered? Who should do the annotating? How should data be validated? Include any acceptable tradeoffs.

- *Name of the requirement. Description.*
- *Name of the requirement. Description.*

Performance requirements

What can people who use this dataset for its intended uses expect?

- *Name of the requirement. Description.*
- *Name of the requirement. Description.*

Maintenance requirements

Should the data be regularly updated? If so, how often? For how long should the data be retained? Include any acceptable tradeoffs.

Sharing requirements

Should the data be made available to other teams within Google and/or open-sourced? If so, what constraints on data licensing, access, usage, and distribution are needed? Include any acceptable tradeoffs.

Caveats and risks

What would be the consequences of using data meeting the requirements described above?

Data ethics

Document your considerations of the ethical implications of the data and its collection.

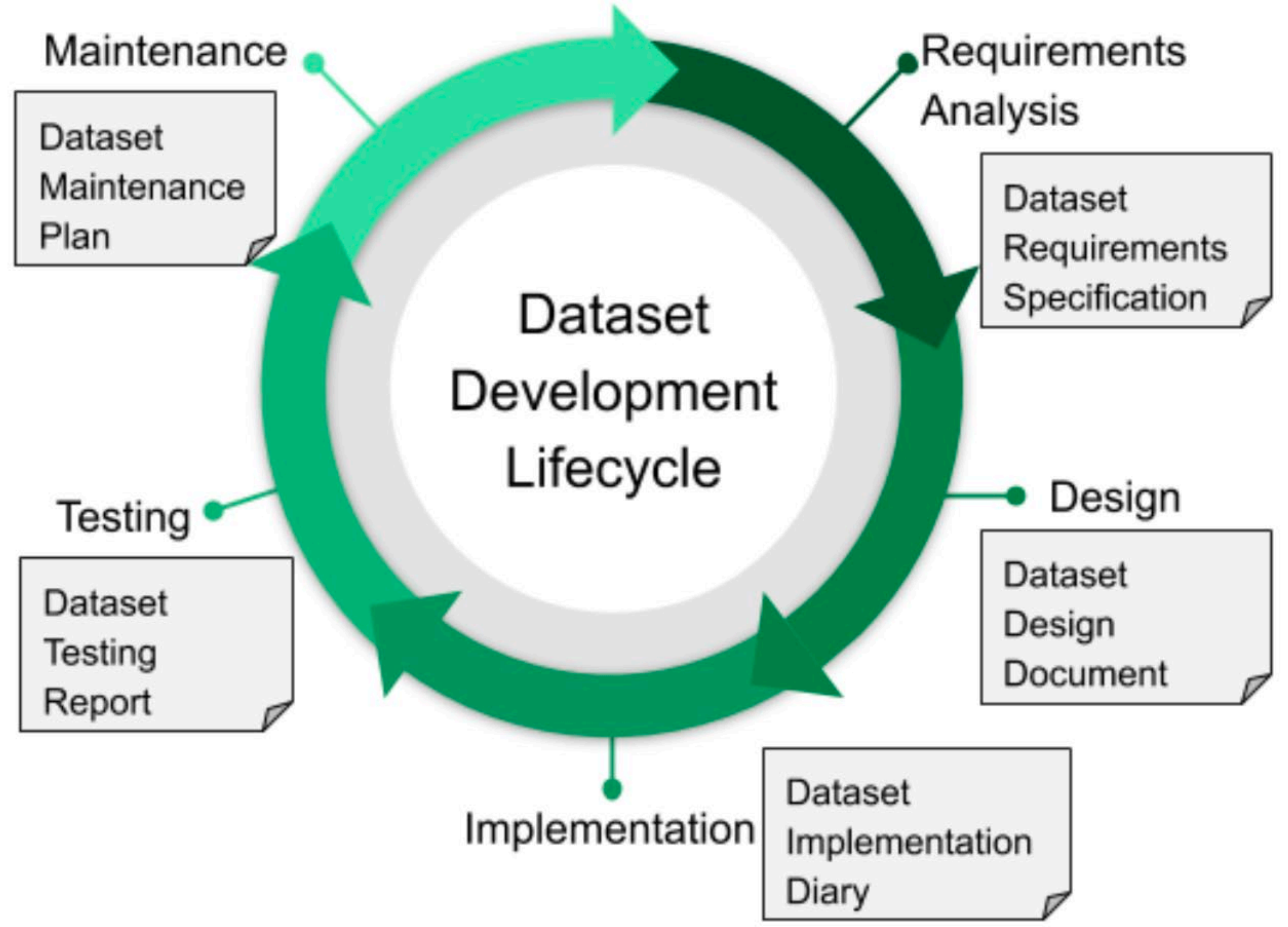


Figure 1: The Dataset Development Lifecycle requires documentation for each stage. See Table 3 for descriptions of each stage, and Table 1 for document types.

Appendix A

The Dataset Nutrition Label:
A Framework To Drive Higher Data Quality Standards

Sarah Holland[1]*, Ahmed Hosny[2]*, Sarah Newman[3], Joshua Joseph[4], and Kasia Chmielinski[1*†]

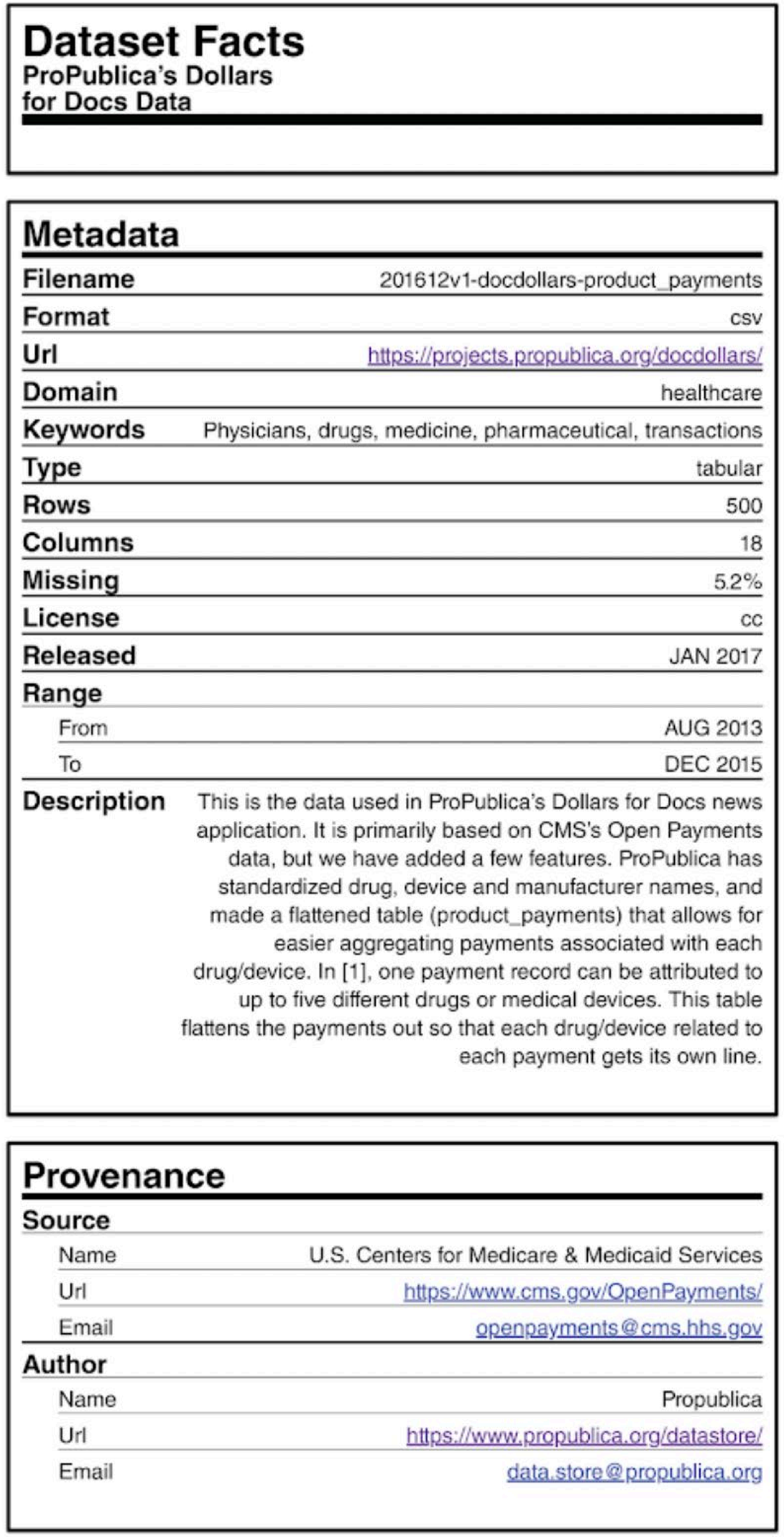
[1]Assembly, MIT Media Lab and Berkman Klein Center at Harvard University, [2]Dana-Farber Cancer Institute, Harvard Medical School, [3]metaLAB (at) Harvard, Berkman Klein Center for Internet & Society, Harvard University, [4]33x.ai

**authors contributed equally*

[†]nutrition@media.mit.edu

Module Name	Description	Contents
Metadata	Meta information. This module is the only required module. It represents the absolute minimum information to be presented	Filename, file format, URL, domain, keywords, type, dataset size, % of missing cells, license, release date, collection range, description
Provenance	Information regarding the origin and lineage of the dataset	Source and author contact information with version history
Variables	Descriptions of each variable (column) in the dataset	Textual descriptions
Statistics	Simple statistics for all variables, in addition to stratifications into ordinal, nominal, continuous, and discrete	Least/most frequent entries, min/max, median, mean,.etc
Pair Plots	Distributions and linear correlations between 2 chosen variables	Histograms and heatmaps
Probabilistic Model	Synthetic data generated using distribution hypotheses from which the data was drawn - leverages a probabilistic programming backend	Histograms and other statistical plots
Ground Truth Correlations	Linear correlations between a chosen variable in the dataset and variables from other datasets considered to be "ground truth", such as Census Data	Heatmaps

Table 1. Table illustrating 7 modules of the Dataset Nutrition Label, together with their description, role, and contents.



Dataset Facts
ProPublica's Dollars for Docs Data

Metadata

Filename	201612v1-docdollars-product_payments
Format	csv
Url	https://projects.propublica.org/docdollars/
Domain	healthcare
Keywords	Physicians, drugs, medicine, pharmaceutical, transactions
Type	tabular
Rows	500
Columns	18
Missing	5.2%
License	cc
Released	JAN 2017
Range	
From	AUG 2013
To	DEC 2015
Description	This is the data used in ProPublica's Dollars for Docs news application. It is primarily based on CMS's Open Payments data, but we have added a few features. ProPublica has standardized drug, device and manufacturer names, and made a flattened table (product_payments) that allows for easier aggregating payments associated with each drug/device. In [1], one payment record can be attributed to up to five different drugs or medical devices. This table flattens the payments out so that each drug/device related to each payment gets its own line.

Provenance

Source	
Name	U.S. Centers for Medicare & Medicaid Services
Url	https://www.cms.gov/OpenPayments/
Email	openpayments@cms.hhs.gov
Author	
Name	Propublica
Url	https://www.propublica.org/datastore/
Email	data.store@propublica.org

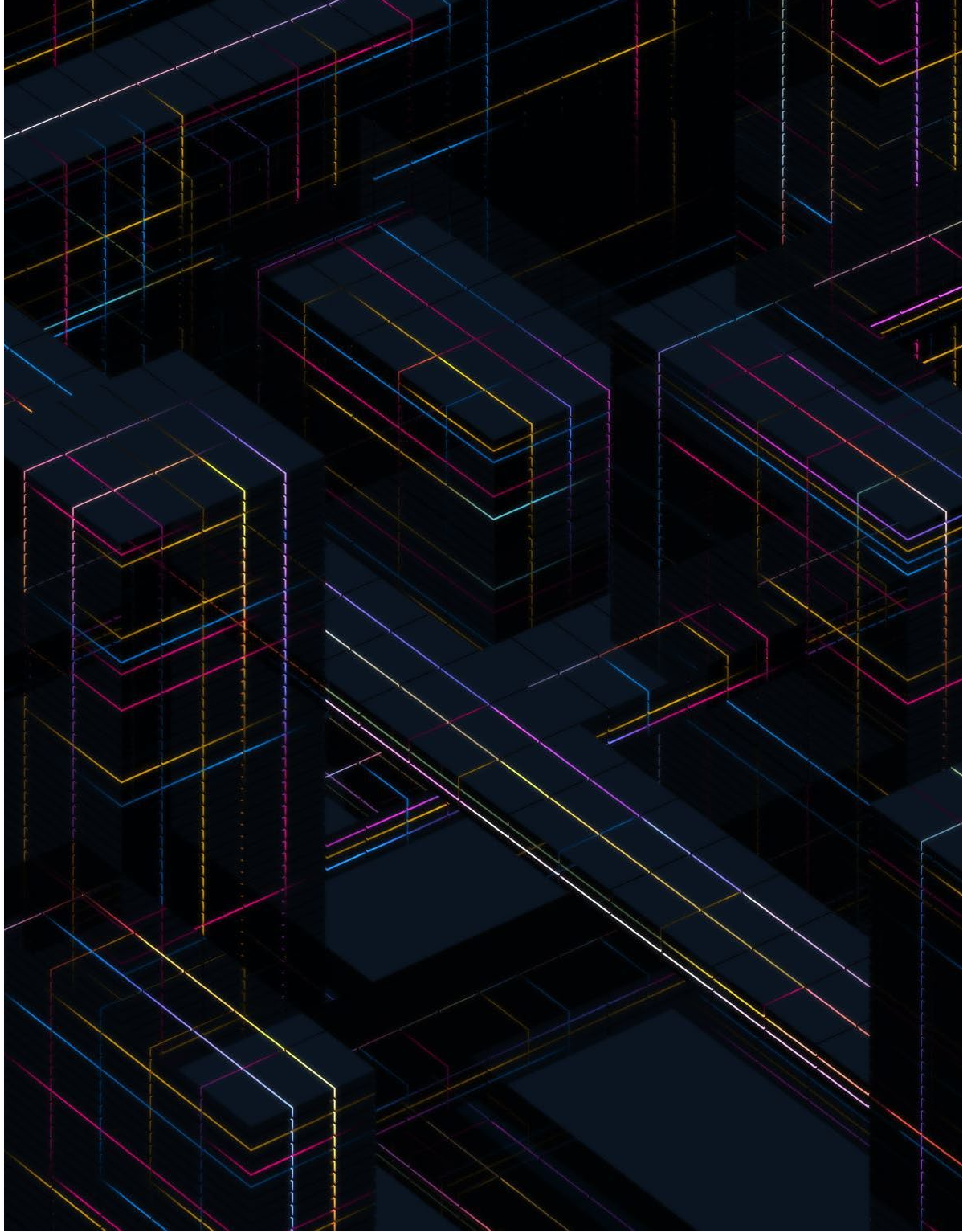
Metadata for AI Models

Tension

- Complexity of ML models give them the ability to learn deeper patterns in data.

- This complexity makes models hard to interpret but most ML is a function of data + ML architecture.

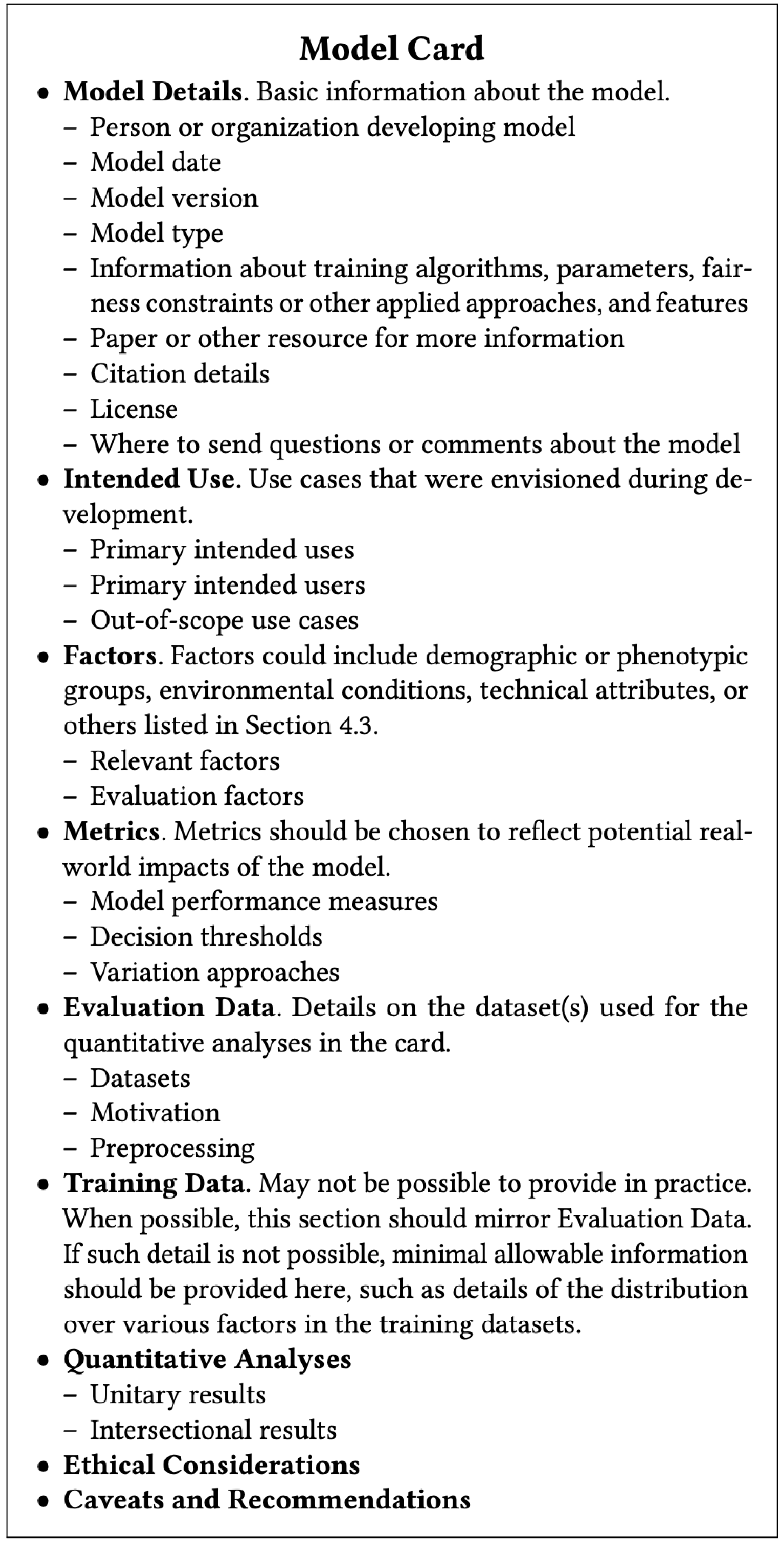
- If the models cannot be transparent, then we need to be transparent about things around the ML models as much as possible.



Model Cards for Model Reporting

Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, Timnit Gebru
{mmitchellai,simonewu,andrewzaldivar,parkerbarnes,lucyvasserman,benhutch,espitzer,tgebru}@google.com
deborah.raji@mail.utoronto.ca

- Stakeholders targeted:
  - ML/AI Practitioners + Developers
  - Policymakers
  - ML-Knowledgeable individuals
  - Impacted individuals



Model Card

- **Model Details**. Basic information about the model.
  - Person or organization developing model
  - Model date
  - Model version
  - Model type
  - Information about training algorithms, parameters, fairness constraints or other applied approaches, and features
  - Paper or other resource for more information
  - Citation details
  - License
  - Where to send questions or comments about the model
- **Intended Use**. Use cases that were envisioned during development.
  - Primary intended uses
  - Primary intended users
  - Out-of-scope use cases
- **Factors**. Factors could include demographic or phenotypic groups, environmental conditions, technical attributes, or others listed in Section 4.3.
  - Relevant factors
  - Evaluation factors
- **Metrics**. Metrics should be chosen to reflect potential real-world impacts of the model.
  - Model performance measures
  - Decision thresholds
  - Variation approaches
- **Evaluation Data**. Details on the dataset(s) used for the quantitative analyses in the card.
  - Datasets
  - Motivation
  - Preprocessing
- **Training Data**. May not be possible to provide in practice. When possible, this section should mirror Evaluation Data. If such detail is not possible, minimal allowable information should be provided here, such as details of the distribution over various factors in the training datasets.
- **Quantitative Analyses**
  - Unitary results
  - Intersectional results
- **Ethical Considerations**
- **Caveats and Recommendations**

Figure 1: Summary of model card sections and suggested prompts for each.

Model Card - Smiling Detection in Images

Model Details

- Developed by researchers at Google and the University of Toronto, 2018, v1.
- Convolutional Neural Net.
- Pretrained for face recognition then fine-tuned with cross-entropy loss for binary smiling classification.

Intended Use

- Intended to be used for fun applications, such as creating cartoon smiles on real images; augmentative applications, such as providing details for people who are blind; or assisting applications such as automatically finding smiling photos.
- Particularly intended for younger audiences.
- Not suitable for emotion detection or determining affect; smiles were annotated based on physical appearance, and not underlying emotions.

Factors

- Based on known problems with computer vision face technology, potential relevant factors include groups for gender, age, race, and Fitzpatrick skin type; hardware factors of camera type and lens type; and environmental factors of lighting and humidity.
- Evaluation factors are gender and age group, as annotated in the publicly available dataset CelebA [36]. Further possible factors not currently available in a public smiling dataset. Gender and age determined by third-party annotators based on visual presentation, following a set of examples of male/female gender and young/old age. Further details available in [36].

Metrics

- Evaluation metrics include **False Positive Rate** and **False Negative Rate** to measure disproportionate model performance errors across subgroups. **False Discovery Rate** and **False Omission Rate**, which measure the fraction of negative (not smiling) and positive (smiling) predictions that are incorrectly predicted to be positive and negative, respectively, are also reported. [48]
- Together, these four metrics provide values for different errors that can be calculated from the confusion matrix for binary classification systems.
- These also correspond to metrics in recent definitions of "fairness" in machine learning (cf. [6, 26]), where parity across subgroups for different metrics correspond to different fairness criteria.
- 95% confidence intervals calculated with bootstrap resampling.
- All metrics reported at the .5 decision threshold, where all error types (FPR, FNR, FDR, FOR) are within the same range (0.04 - 0.14).

Training Data

- CelebA [36], training data split.

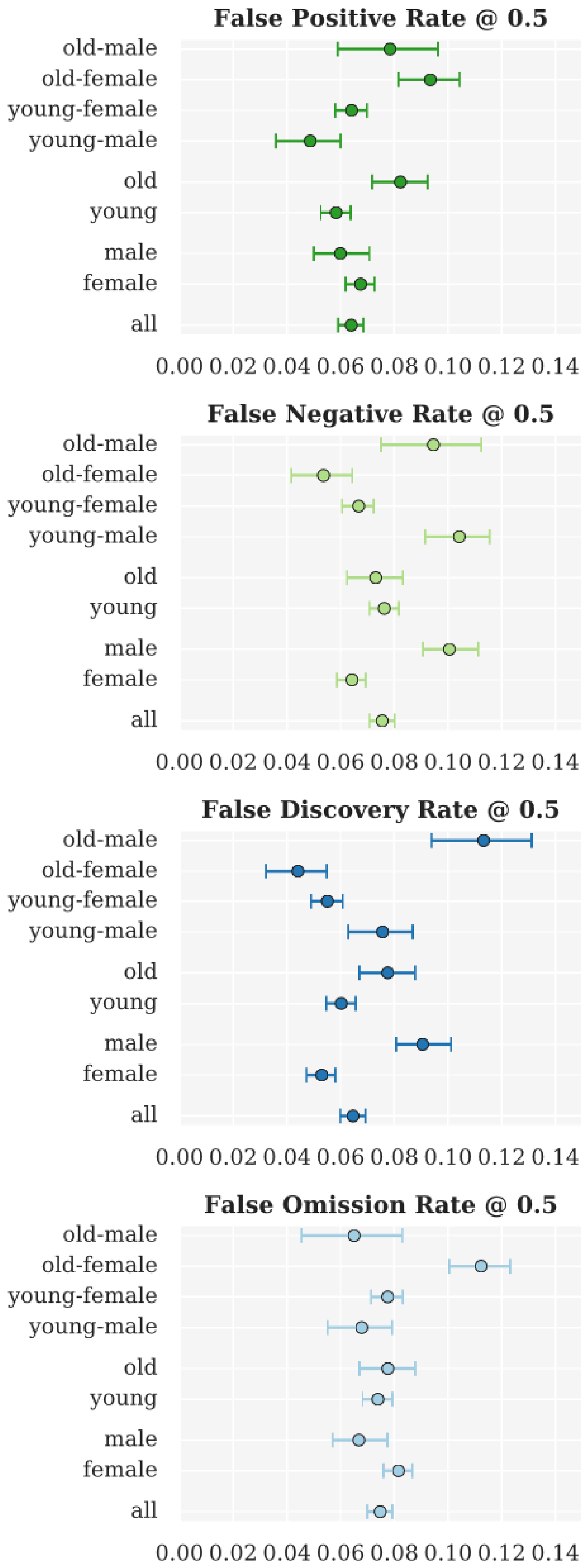
Evaluation Data

- CelebA [36], test data split.
- Chosen as a basic proof-of-concept.

Ethical Considerations

- Faces and annotations based on public figures (celebrities). No new information is inferred or annotated.

Quantitative Analyses



Caveats and Recommendations

- Does not capture race or skin type, which has been reported as a source of disproportionate errors [5].
- Given gender classes are binary (male/not male), which we include as male/female. Further work needed to evaluate across a spectrum of genders.
- An ideal evaluation dataset would additionally include annotations for Fitzpatrick skin type, camera details, and environment (lighting/humidity) details.

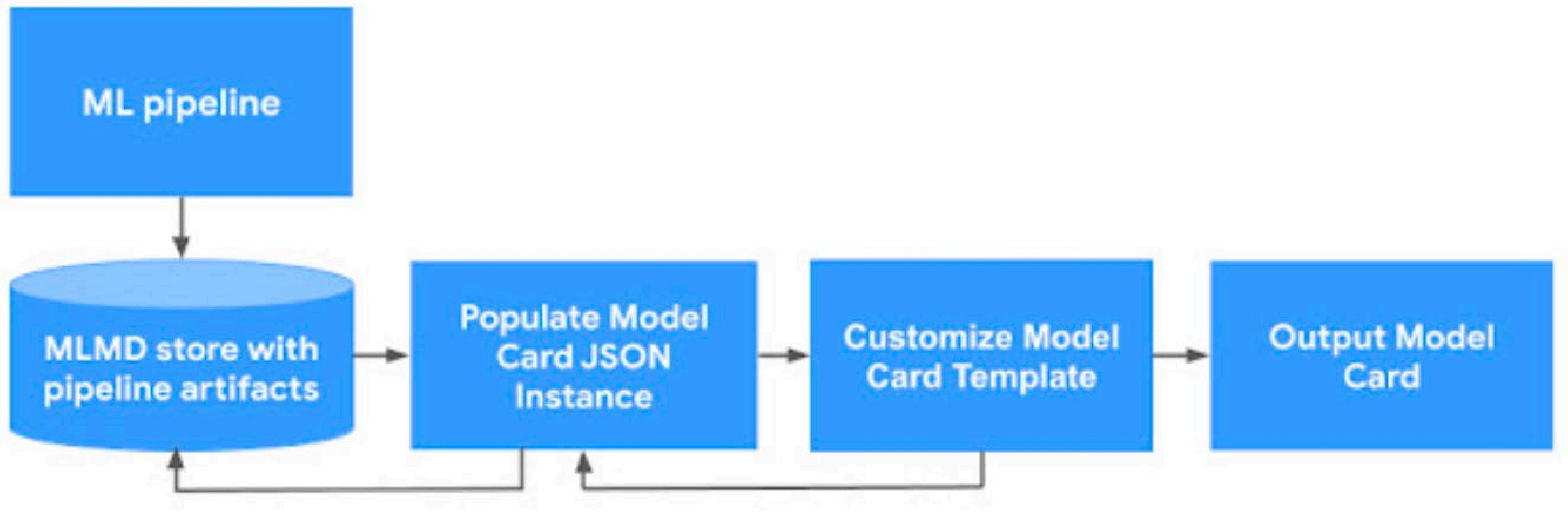
Google Research   Philosophy   Research Areas   Publications   People   Resources

BLOG ›

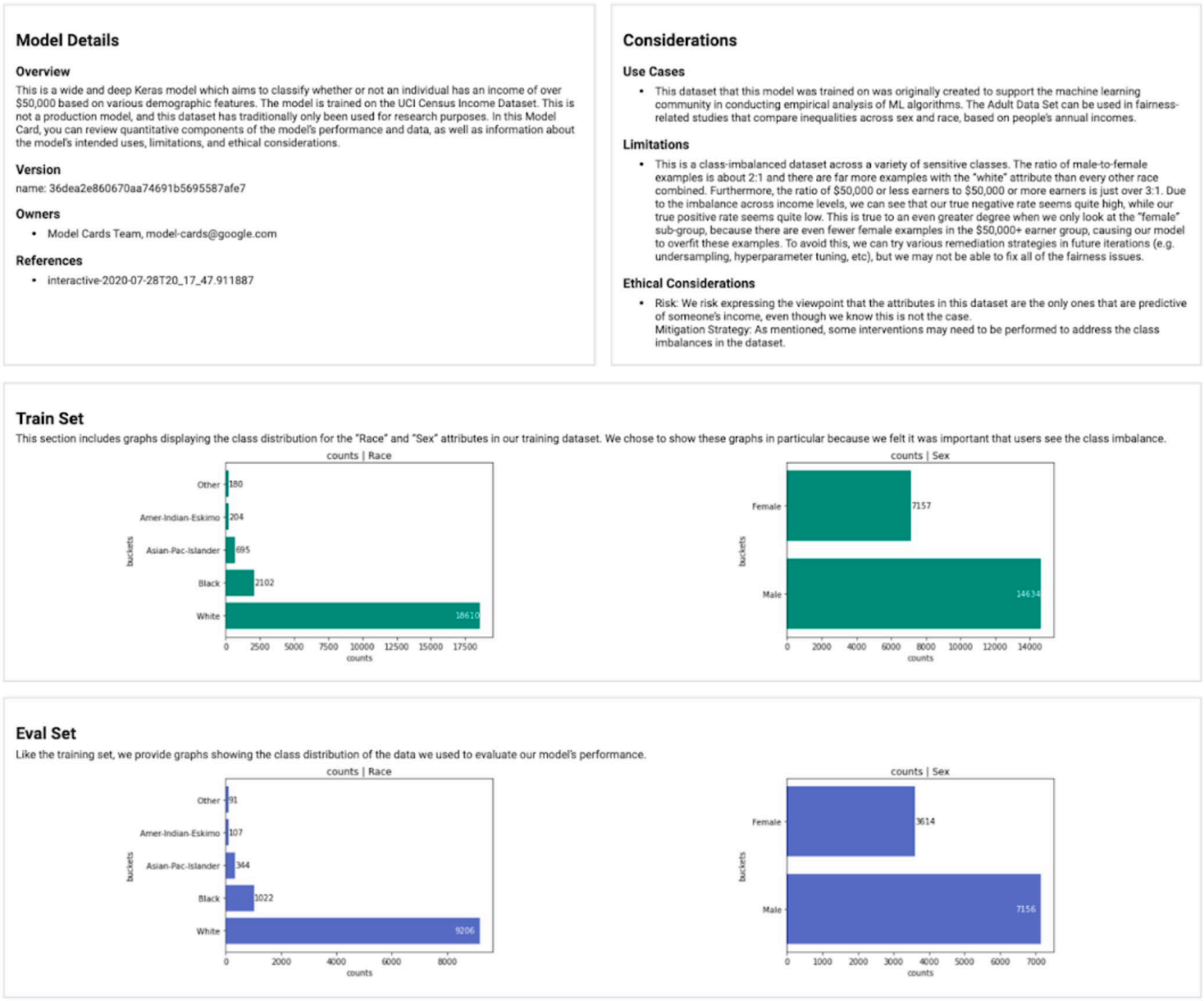
Introducing the Model Card Toolkit for Easier Model Transparency Reporting

ÇARŞAMBA, TEMMUZ 29, 2020

Posted by Huanming Fang and Hui Miao, Software Engineers, Google Research



ML pipeline
MLMD store with pipeline artifacts
Populate Model Card JSON Instance
Customize Model Card Template
Output Model Card



Model Details

Overview

This is a wide and deep Keras model which aims to classify whether or not an individual has an income of over $50,000 based on various demographic features. The model is trained on the UCI Census Income Dataset. This is not a production model, and this dataset has traditionally only been used for research purposes. In this Model Card, you can review quantitative components of the model's performance and data, as well as information about the model's intended uses, limitations, and ethical considerations.

Version

name: 36dea2e860670aa74691b5695587afe7

Owners

- Model Cards Team, model-cards@google.com

References

- interactive-2020-07-28T20_17_47.911887

Considerations

Use Cases

- This dataset that this model was trained on was originally created to support the machine learning community in conducting empirical analysis of ML algorithms. The Adult Data Set can be used in fairness-related studies that compare inequalities across sex and race, based on people's annual incomes.

Limitations

- This is a class-imbalanced dataset across a variety of sensitive classes. The ratio of male-to-female examples is about 2:1 and there are far more examples with the "white" attribute than every other race combined. Furthermore, the ratio of $50,000 or less earners to $50,000 or more earners is just over 3:1. Due to the imbalance across income levels, we can see that our true negative rate seems quite high, while our true positive rate seems quite low. This is true to an even greater degree when we only look at the "female" sub-group, because there are even fewer female examples in the $50,000+ earner group, causing our model to overfit these examples. To avoid this, we can try various remediation strategies in future iterations (e.g. undersampling, hyperparameter tuning, etc), but we may not be able to fix all of the fairness issues.

Ethical Considerations

- Risk: We risk expressing the viewpoint that the attributes in this dataset are the only ones that are predictive of someone's income, even though we know this is not the case.
  Mitigation Strategy: As mentioned, some interventions may need to be performed to address the class imbalances in the dataset.

Train Set

This section includes graphs displaying the class distribution for the "Race" and "Sex" attributes in our training dataset. We chose to show these graphs in particular because we felt it was important that users see the class imbalance.

Eval Set

Like the training set, we provide graphs showing the class distribution of the data we used to evaluate our model's performance.

🤗 Hugging Face

- An AI startup originally focused on making a chatbot for teens.
- Pivoted towards trying to build a community and ecosystem of tools for accelerating AI research
- Started mainly in the NLP space.
- Provided easy to use interfaces to Text-based DL models (transformers models that worked with TF/PyTorch)
- Evolved to include LLMs

THE LANDSCAPE OF ML DOCUMENTATION TOOLS

The development of the model cards framework in 2018 was inspired by the major documentation framework efforts of Data Statements for Natural Language Processing (Bender & Friedman, 2018) and Datasheets for Datasets (Gebru et al., 2018). Since model cards were proposed, a number of other tools have been proposed for documenting and evaluating various aspects of the machine learning development cycle. These tools, including model cards and related documentation efforts proposed prior to model cards, can be contextualised with regard to their focus (e.g., on which part of the ML system lifecycle does the tool focus?) and their intended audiences (e.g., who is the tool designed for?). In Figures 1-2 below, we summarise several prominent documentation tools along these dimensions, provide contextual descriptions of each tool, and link to examples. We broadly classify the documentation tools as belong to the following groups:

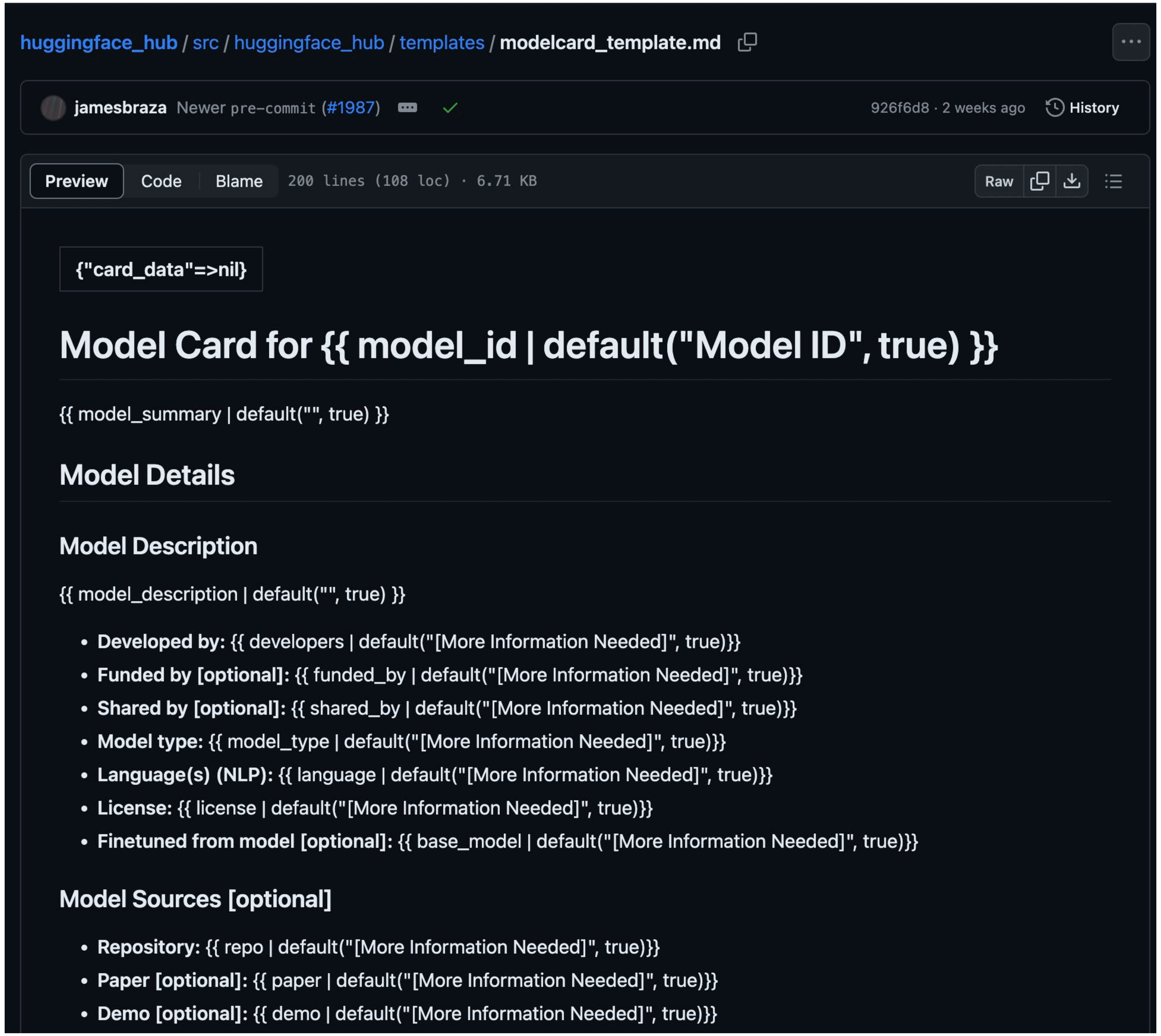
- **Data-focused**, including documentation tools focused on datasets used in the machine learning system lifecycle
- **Models-and-methods-focused**, including documentation tools focused on machine learning models and methods; and
- **Systems-focused**, including documentation tools focused on ML systems, including models, methods, datasets, APIs, and non AI/ML components that interact with each other as part of an ML system

User Study Details

We selected people from a variety of different backgrounds relevant to machine learning and model documentation. Below, we detail their demographics, the questions they were asked, and the corresponding insights from their responses. Full details on responses are available in Appendix A.

Respondent Demographics

- Tech & Regulatory Affairs Counsel
- ML Engineer (x2)
- Developer Advocate
- Executive Assistant
- Monetization Lead
- Policy Manager/AI Researcher
- Research Intern


huggingface_hub / src / huggingface_hub / templates / modelcard_template.md
jamesbraza Newer pre-commit (#1987)
926f6d8 · 2 weeks ago
History
Preview
Code
Blame
200 lines (108 loc) · 6.71 KB
Raw

{"card_data"=>nil}

Model Card for {{ model_id | default("Model ID", true) }}

{{ model_summary | default("", true) }}

Model Details

Model Description

{{ model_description | default("", true) }}

- **Developed by:** {{ developers | default("[More Information Needed]", true)}}
- **Funded by [optional]:** {{ funded_by | default("[More Information Needed]", true)}}
- **Shared by [optional]:** {{ shared_by | default("[More Information Needed]", true)}}
- **Model type:** {{ model_type | default("[More Information Needed]", true)}}
- **Language(s) (NLP):** {{ language | default("[More Information Needed]", true)}}
- **License:** {{ license | default("[More Information Needed]", true)}}
- **Finetuned from model [optional]:** {{ base_model | default("[More Information Needed]", true)}}

Model Sources [optional]

- **Repository:** {{ repo | default("[More Information Needed]", true)}}
- **Paper [optional]:** {{ paper | default("[More Information Needed]", true)}}
- **Demo [optional]:** {{ demo | default("[More Information Needed]", true)}}

Template

modelcard_template.md file

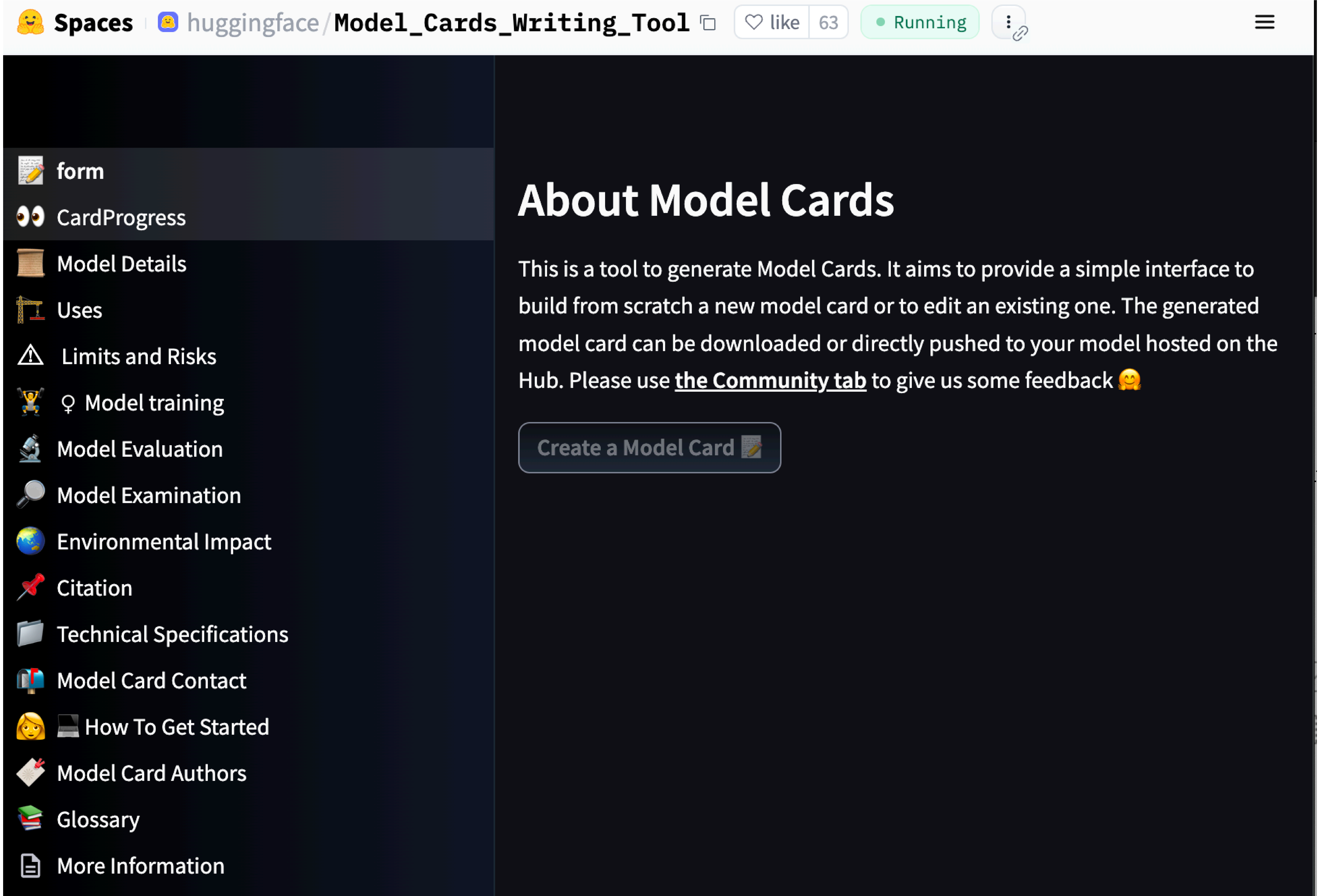
Directions

Fully filling out a model card requires input from a few different roles. (One person may have more than one role.) We'll refer to these roles as the **developer**, who writes the code and runs training; the **sociotechnic**, who is skilled at analyzing the interaction of technology and society long-term (this includes lawyers, ethicists, sociologists, or rights advocates); and the **project organizer**, who understands the overall scope and reach of the model, can roughly fill out each part of the card, and who serves as a contact person for model card updates.

- The **developer** is necessary for filling out Training Procedure and Technical Specifications. They are also particularly useful for the "Limitations" section of Bias, Risks, and Limitations. They are responsible for providing Results for the Evaluation, and ideally work with the other roles to define the rest of the Evaluation: Testing Data, Factors & Metrics.
- The **sociotechnic** is necessary for filling out "Bias" and "Risks" within Bias, Risks, and Limitations, and particularly useful for "Out of Scope Use" within Uses.
- The **project organizer** is necessary for filling out Model Details and Uses. They might also fill out Training Data. Project organizers could also be in charge of Citation, Glossary, Model Card Contact, Model Card Authors, and More Information.

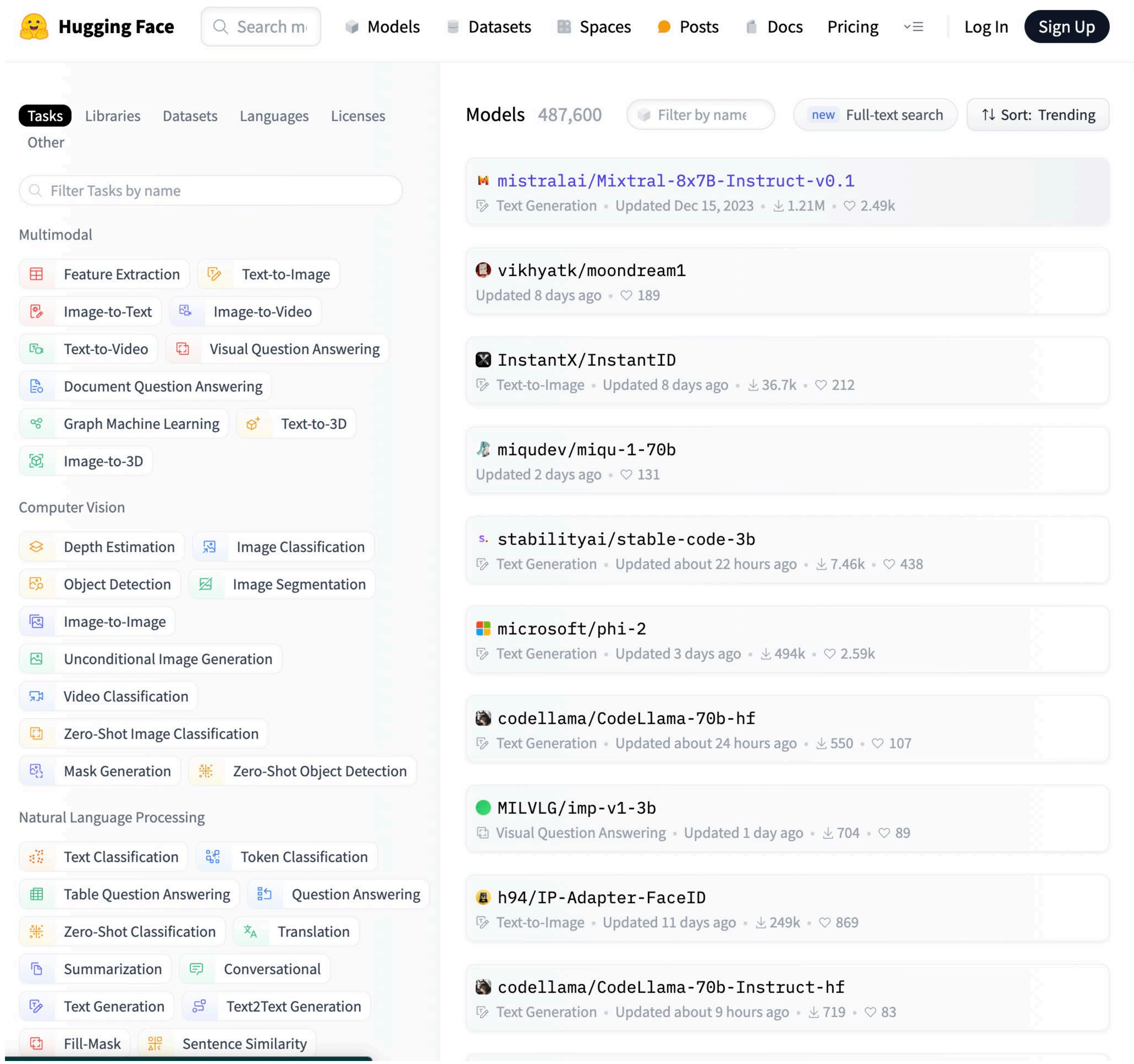
Instructions are provided below, in italics.

Template variable names appear in `monospace`.



About Model Cards

This is a tool to generate Model Cards. It aims to provide a simple interface to build from scratch a new model card or to edit an existing one. The generated model card can be downloaded or directly pushed to your model hosted on the Hub. Please use **the Community tab** to give us some feedback 🤗



Multimodal

Feature Extraction · Text-to-Image · Image-to-Text · Image-to-Video · Text-to-Video · Visual Question Answering · Document Question Answering · Graph Machine Learning · Text-to-3D · Image-to-3D

Computer Vision

Depth Estimation · Image Classification · Object Detection · Image Segmentation · Image-to-Image · Unconditional Image Generation · Video Classification · Zero-Shot Image Classification · Mask Generation · Zero-Shot Object Detection

Natural Language Processing

Text Classification · Token Classification · Table Question Answering · Question Answering · Zero-Shot Classification · Translation · Summarization · Conversational · Text Generation · Text2Text Generation · Fill-Mask · Sentence Similarity

Models 487,600

mistralai/Mixtral-8x7B-Instruct-v0.1
Text Generation • Updated Dec 15, 2023 • 1.21M • ♡ 2.49k

vikhyatk/moondream1
Updated 8 days ago • ♡ 189

InstantX/InstantID
Text-to-Image • Updated 8 days ago • 36.7k • ♡ 212

miqudev/miqu-1-70b
Updated 2 days ago • ♡ 131

stabilityai/stable-code-3b
Text Generation • Updated about 22 hours ago • 7.46k • ♡ 438

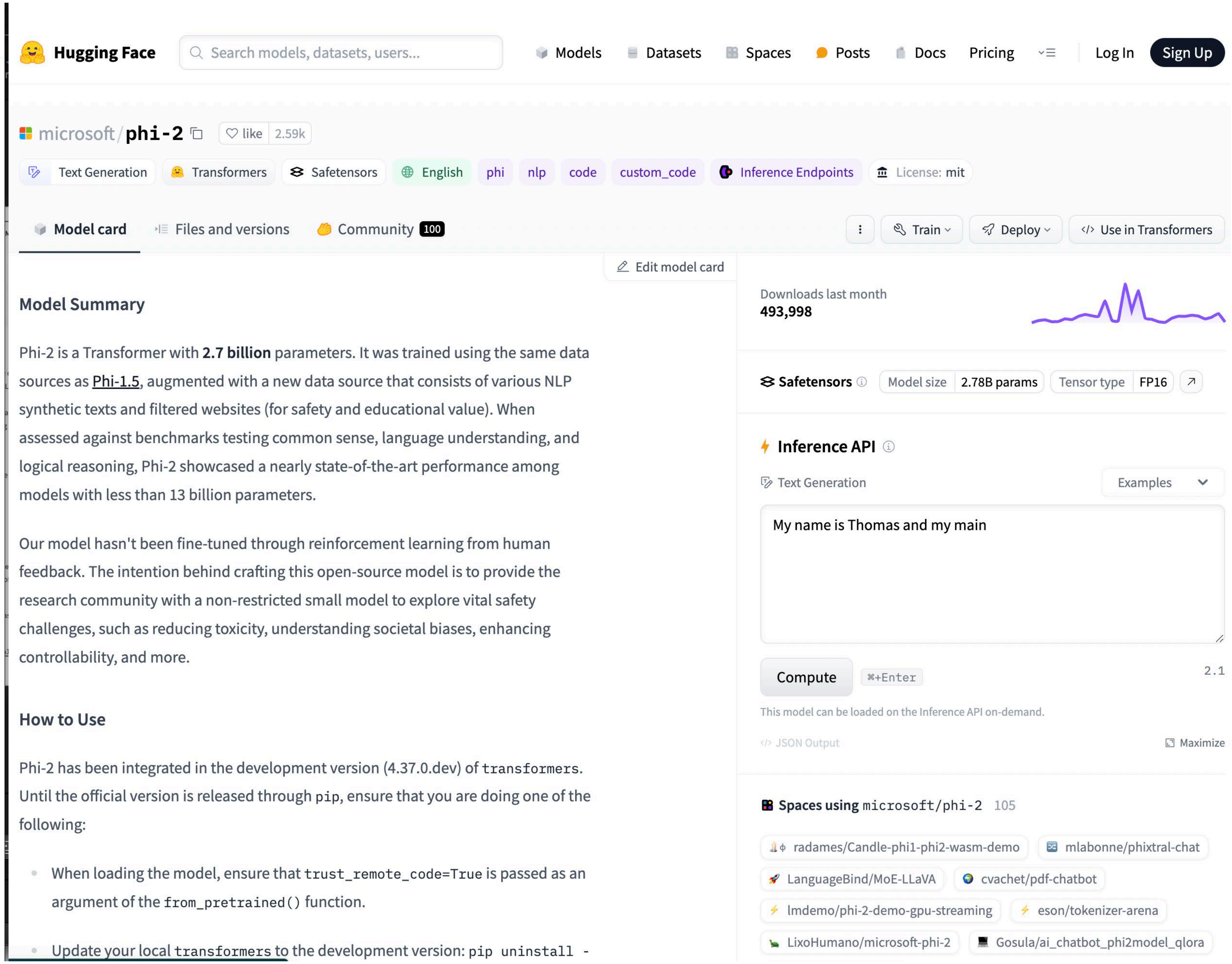
microsoft/phi-2
Text Generation • Updated 3 days ago • 494k • ♡ 2.59k

codellama/CodeLlama-70b-hf
Text Generation • Updated about 24 hours ago • 550 • ♡ 107

MILVLG/imp-v1-3b
Visual Question Answering • Updated 1 day ago • 704 • ♡ 89

h94/IP-Adapter-FaceID
Text-to-Image • Updated 11 days ago • 249k • ♡ 869

codellama/CodeLlama-70b-Instruct-hf
Text Generation • Updated about 9 hours ago • 719 • ♡ 83



microsoft/phi-2

Text Generation · Transformers · Safetensors · English · phi · nlp · code · custom_code · Inference Endpoints · License: mit

Model Summary

Phi-2 is a Transformer with **2.7 billion** parameters. It was trained using the same data sources as Phi-1.5, augmented with a new data source that consists of various NLP synthetic texts and filtered websites (for safety and educational value). When assessed against benchmarks testing common sense, language understanding, and logical reasoning, Phi-2 showcased a nearly state-of-the-art performance among models with less than 13 billion parameters.

Our model hasn't been fine-tuned through reinforcement learning from human feedback. The intention behind crafting this open-source model is to provide the research community with a non-restricted small model to explore vital safety challenges, such as reducing toxicity, understanding societal biases, enhancing controllability, and more.

How to Use

Phi-2 has been integrated in the development version (4.37.0.dev) of `transformers`. Until the official version is released through `pip`, ensure that you are doing one of the following:

- When loading the model, ensure that `trust_remote_code=True` is passed as an argument of the `from_pretrained()` function.
- Update your local `transformers` to the development version: `pip uninstall -`

Downloads last month
493,998

Safetensors · Model size 2.78B params · Tensor type FP16

Inference API

Text Generation

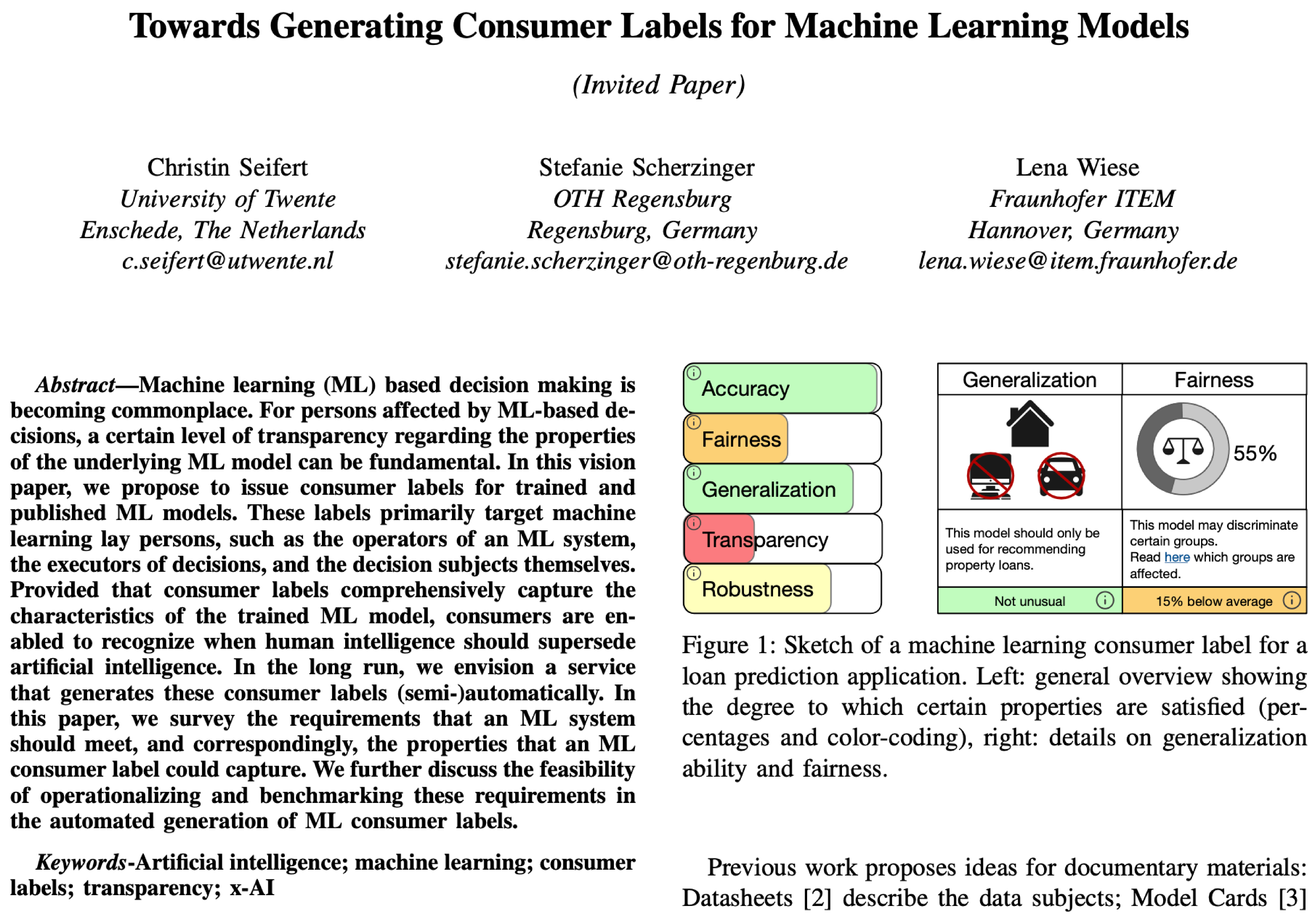
My name is Thomas and my main

2.1

This model can be loaded on the Inference API on-demand.

Spaces using `microsoft/phi-2` 105

radames/Candle-phi1-phi2-wasm-demo · mlabonne/phixtral-chat · LanguageBind/MoE-LLaVA · cvachet/pdf-chatbot · lmdemo/phi-2-demo-gpu-streaming · eson/tokenizer-arena · LixoHumano/microsoft-phi-2 · Gosula/ai_chatbot_phi2model_qlora



Towards Generating Consumer Labels for Machine Learning Models

(Invited Paper)

Christin Seifert
University of Twente
Enschede, The Netherlands
c.seifert@utwente.nl

Stefanie Scherzinger
OTH Regensburg
Regensburg, Germany
stefanie.scherzinger@oth-regenburg.de

Lena Wiese
Fraunhofer ITEM
Hannover, Germany
lena.wiese@item.fraunhofer.de

***Abstract*—Machine learning (ML) based decision making is becoming commonplace. For persons affected by ML-based decisions, a certain level of transparency regarding the properties of the underlying ML model can be fundamental. In this vision paper, we propose to issue consumer labels for trained and published ML models. These labels primarily target machine learning lay persons, such as the operators of an ML system, the executors of decisions, and the decision subjects themselves. Provided that consumer labels comprehensively capture the characteristics of the trained ML model, consumers are enabled to recognize when human intelligence should supersede artificial intelligence. In the long run, we envision a service that generates these consumer labels (semi-)automatically. In this paper, we survey the requirements that an ML system should meet, and correspondingly, the properties that an ML consumer label could capture. We further discuss the feasibility of operationalizing and benchmarking these requirements in the automated generation of ML consumer labels.**

***Keywords*-Artificial intelligence; machine learning; consumer labels; transparency; x-AI**

Accuracy
Fairness
Generalization
Transparency
Robustness
Generalization
Fairness
55%
This model should only be used for recommending property loans.
This model may discriminate certain groups. Read here which groups are affected.
Not unusual
15% below average

Figure 1: Sketch of a machine learning consumer label for a loan prediction application. Left: general overview showing the degree to which certain properties are satisfied (percentages and color-coding), right: details on generalization ability and fairness.

Previous work proposes ideas for documentary materials: Datasheets [2] describe the data subjects; Model Cards [3]

● DATA FOCUSED

- Data Sheets ● ● ● ● ● ● ●
- Data Statements ● ● ● ● ●
- Data Nutrition Labels ● ● ● ● ●
- Data Cards for NLP ● ● ● ● ●
- Dataset Development Lifecycle Documentation Framework ● ● ● ● ● ● ●
- Data Cards ● ● ● ● ●

● MODELS & METHODS FOCUSED

- Model Cards ● ● ● ● ● ● ●
- Value Cards ● ● ● ● ●
- Method Cards ● ●
- Consumer Labels for Models ● ● ● ●

● SYSTEMS FOCUSED

- System Cards ● ● ● ● ●
- FactSheets ● ● ● ●
- ABOUT ML ● ● ● ● ● ● ●

SAMPLE OF POTENTIAL AUDIENCES

● ML Engineers
● Model Developers/Reviewers
● Students
● Policymakers
● Ethicists
● Data Scientists/Business Analysts
● Impacted Individuals

https://huggingface.co/blog/model-cards

Team

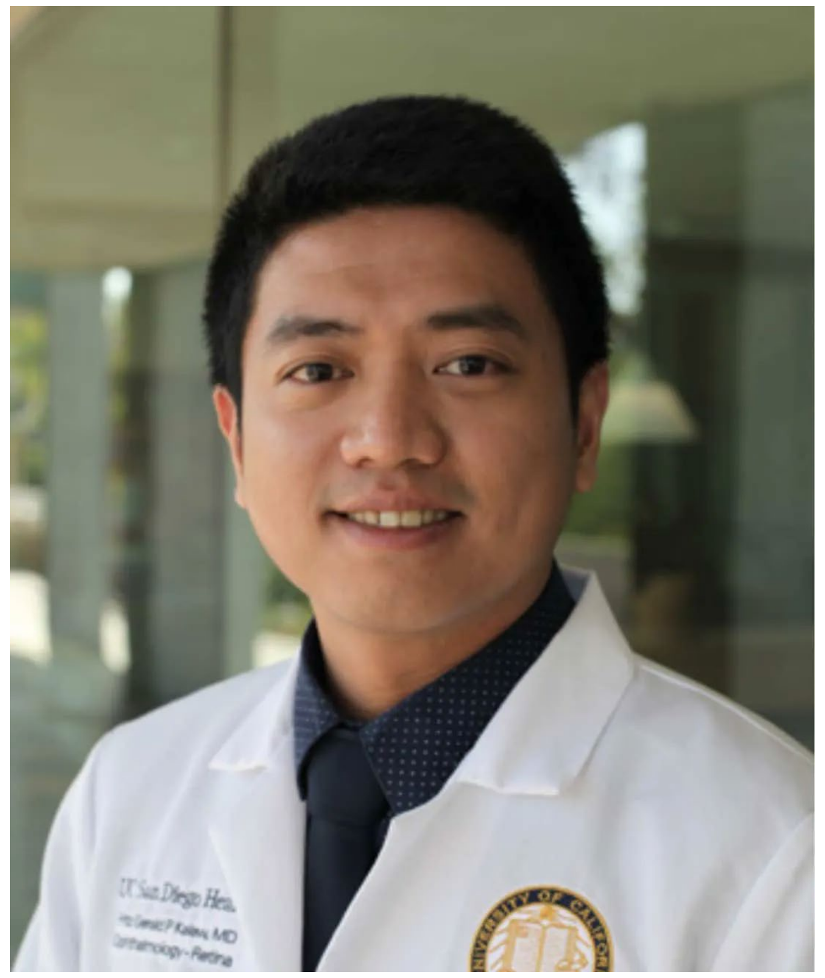
Bhavesh Patel

AI-READI

AI Ready and Equitable Atlas for Diabetes Insights

Anna Heinke

Lingling Huang



Fritz Kalaw

Kyongmi Simpkins

UNIVERSITY of WASHINGTON

Acknowledgements

UW Medicine
DEPARTMENT OF OPHTHALMOLOGY

Funding Sources: NIH 1OT2OD032644, NIH/NIA R01 AG060942, NIH/NEI K23EY024921, NIA/NIH U19AG066567, Research to Prevent Blindness, Latham Vision Research Innovation Award, Latham Fund for Vision Research, Donors to Computational Ophthalmology Fund

THE LOWY MEDICAL RESEARCH INSTITUTE

NIH National Institute on Aging

NIH National Eye Institute

Research to Prevent Blindness

Aaron Lee, MD MSCI

Cecilia Lee, MD MS

Computational Ophthalmology

https://comp.ophthalmology.uw.edu

Megan Lacy MS
Julia Owen PhD
Yue Wu PhD
Scott Song BA

Missy Takahashi BS
Ashley Batchelor MS
Matthew Hunt BS
Theodore Spaide PhD

Emily Heindsmann MA
Christina Duong BS COA
Yelena Bagdasarova PhD

Marian Blazes MD
Jamie Shaffer MS
Yuka Kihara MS
Randy Lu BS